



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 201729**

**TO: Manjunath N Rao**  
**Location: REM/2A01/2C70**  
**Art Unit: 1652**  
**September 19, 2006**  
  
**Case Serial Number: 10/815774**

**From: P. Sheppard**  
**Location: Remsen Building**  
**Phone: (571) 272-2529**

**sheppard@uspto.gov**

### **Search Notes**

**From:** Chan, Christina  
**Sent:** Thursday, September 14, 2006 3:56 PM  
**To:** Rao, Manjunath N.; STIC-Biotech/ChemLib  
**Subject:** RE: RUSH Sequence search for 10/815774

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

**From:** Rao, Manjunath N.  
**Sent:** Thursday, September 14, 2006 3:43 PM  
**To:** Chan, Christina  
**Subject:** RUSH Sequence search for 10/815774

Hello Christina

Please authorize the search request below as RUSH. The reason being this is a date goal case and claims are drawn to sequences.

Many Thanks  
-Manjunath Rao

Manjunath N. Rao  
Primary Examiner  
Art Unit 1652  
Ph: 571-272-0939

---

**From:** Manjunath N. Rao  
**Art Unit 1652, Room 2A01**  
**Mail Box in Room 2C70**  
Phone: 272-0939

**Date:** 9-14-06

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIS: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

Please search the following as soon as possible for application with serial number  
**10/815,774**

1. SEQ ID NO: 5 against all commercial nucleic acid sequence databases, issued patents/published applications nucleic acid sequence database and pending application nucleic acid sequence database. Please provide a print of results
2. SEQ ID NO: 6 against all commercial amino acid sequence databases, issued patents/published applications amino acid sequence database and pending application amino acid sequence database. Please provide a print of results.

If you have any questions please call me at the above phone number.

Thanks

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

This Page Blank (uspto)

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2006, 14:47:34 / Search time 3717 Seconds  
(without alignments)  
9548.243 Million cell updates/sec

Title: US-10-815-774-5

Perfect score: 555  
Sequence: 1 ttggcgagtgtaattcctga.....atgtaccagctgtgattc 555

Scoring table: IDENTITY NJC  
Gapop 10.0, Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

GenEmbl:\*  
1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_sts:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vi:\*  
11: gb\_ov:\*  
12: gb\_hlg:\*  
13: gb\_in:\*  
14: gb\_om:\*  
15: gb\_ba:\*

Pred. No. is the number of results predicted by chance a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	555	BD015599	Novel pro
2	555	100.0	555	AR560977	Sequence
3	555	100.0	555	AX113621	Sequence
4	555	100.0	1080	BD015600	Novel pro
5	555	100.0	1080	AR560978	Sequence
6	555	100.0	1080	AX113623	Sequence
7	555	100.0	1380	AB046594	Chryseoba
8	344.2	62.0	555	AR159957	Sequence
9	344.2	62.0	555	ES9337	Novel prote
10	344.2	62.0	555	AR570507	Sequence
11	344.2	62.0	1080	AR159962	Sequence
12	344.2	62.0	1080	ES9342	Novel prote
13	344.2	62.0	1080	AR570512	Sequence
14	43	7.7	239553	12 AC098659	AR570512 Sequence
15	41	7.4	125020	5 AC098659	Rattus no
16	40.8	7.4	168540	5 AF429315	Homo sapi
17	39.8	7.2	152202	12 AC129044	Rattus no
18	39.8	7.2	231746	5 AC094081	Homo sapi
				AC011389	Homo sapi

C 19	39.6	7.1	110000	12	CR385089_0	CR385089	Danio rer
C 20	39.6	7.1	110000	12	CR385985_1	CR385985	Continuation (2 of
C 21	39.6	7.1	197752	12	CR387933	CR387933	Danio rer
C 22	39.6	7.1	264742	12	AC162403	AC162403	Bos tauru
C 23	39.4	7.1	239600	12	AC168301	AC168301	Bos tauru
C 24	38.8	7.0	154058	14	AC150680	AC150680	Monodelph
C 25	38.8	7.0	191715	12	AC178434	AC178434	Strongylo
C 26	38.6	7.0	155292	12	AC177955	AC177955	Strongylo
C 27	38.4	6.9	91275	13	AC152924	AC152924	Tribolium
C 28	38.4	6.9	192706	6	AC154283	AC154283	Mus muscu
C 29	38.2	6.9	202982	11	CR788849	CR788849	Zebrafish
C 30	38	6.8	27068	13	CE15A7	CE15A7	Zebrafish
C 31	38	6.8	110000	15	EX908798_00	EX908798	Caenorhabdi
C 32	38	6.8	138682	11	EX908778	EX908778	Parachlam
C 33	38	6.8	178077	11	AL929090	AL929090	Zebrafish
C 34	38	6.8	185947	5	AC092630	AC092630	Zebrafish
C 35	38	6.8	245721	12	AC172971	AC172971	Homo sapi
C 36	37.8	6.8	110000	12	EX897735_1	EX897735	Bos tauru
C 37	37.8	6.8	110000	12	EX897735_2	EX897735	Continuation (2 of
C 38	37.8	6.8	110017	12	CT573372	CT573372	Danio rer
C 39	37.8	6.8	147692	12	CT574555	CT574555	Danio rer
C 40	37.8	6.8	161731	11	AL590155	AL590155	Zebrafish
C 41	37.8	6.8	173576	12	AC166881	AC166881	Oryctolag
C 42	37.8	6.8	219971	6	AL954725	AL954725	Mus muscu
C 43	37.6	6.8	157430	6	AC135378	AC135378	Mus muscu
C 44	37.6	6.8	161223	11	AC022390	AC022390	Homo sapi
C 45	37.6	6.8	184556	6	AC117230	AC117230	Mus muscu

#### ALIGNMENTS

RESULT 1	BD015599	555 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD015599				
DEFINITION	Novel protein deamidase, microorganism producing the same, gene encoding the same, process for producing the same and use of the same.				
ACCESSION	BD015599				
VERSION	BD015599.1	GI:22556736			
KEYWORDS	JP 2001218590-A/3.				
SOURCE	Chryseobacterium sp. No. 9670				
ORGANISM	Chryseobacterium sp. No. 9670				
REFERENCE	1 (bases 1 to 555)				
AUTHORS	Yamaguchi, S.				
TITLE	Novel protein deamidase, microorganism producing the same, gene encoding the same, process for producing the same and use of the same.				
JOURNAL	Patent: JP 2001218590-A 3 14-AUG-2001;				
COMMENT	AMANO ENZYME CO LTD				
OS	Chryseobacterium sp. No. 9670				
PN	JP 2001218590-A/3				
PD	14-AUG-2001				
PF	04-DEC-2000 JP 2000368983				
PI	SHOTARO YAMAGUCHI				
PC	CHRYSEOBACTERIUM SP. NO. 9670				
PC	21, C12N5/10, C07K1/113, C12N1/15, C12N1/19, C12N1/20, C12N1/21, C12N5/00, C12N9/80, C12N1/01, C12N9/80, C12R1/01, C12N15/00, C12N5/00				
CC	Novel protein deamidase, microorganism producing the same, gene encoding				
CC	the same, process for producing the same and use of the same				
FT	Key				
FT	Location/Qualifiers				
FT	1..555				
FEATURES	Location/Qualifiers				
source	1..555				
	/organism="Chryseobacterium sp. No. 9670"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:161487"				
ORIGIN					

Query Match 100.0%; Score 555; DB 2; Length 555;  
Best Local Similarity 100.0%; Pred. No. 3.1e-167;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGAAAT 60  
DB 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGAAAT 60

QY 61 CAGCTTTGGGATCCTCTACAGGCGCTCTCCACATGCATCATCTTCAATTCCTGTAGAC 120  
DB 61 CAGCTTTGGGATCCTCTACAGGCGCTCTCCACATGCATCATCTTCAATTCCTGTAGAC 120

QY 121 GGATGTATGCAAGAGCCCAATAGATGAGCAATCTTAATGAACAACGGCTATGACTGT 180  
DB 121 GGATGTATGCAAGAGCCCAATAGATGAGCAATCTTAATGAACAACGGCTATGACTGT 180

QY 181 GAAATCAATTTGTATACGGAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 240  
DB 181 GAAATCAATTTGTATACGGAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 240

QY 241 AGCTACCAAGTTCATATTTGTAGTATTAATTAATGCTCCGGAGTAAAGGAAAAAGA 300  
DB 241 AGCTACCAAGTTCATATTTGTAGTATTAATTAATGCTCCGGAGTAAAGGAAAAAGA 300

QY 301 ATTATGATCCTTCACATATTTTCAAGCGGCTCTGTAAACAGATACAGATGAGAAACGCT 360  
DB 301 ATTATGATCCTTCACATATTTTCAAGCGGCTCTGTAAACAGATACAGATGAGAAACGCT 360

QY 361 TGCCTTAACACCTCTTGCGGATTCGATCCGTTTCCTTTATGCTAATTAATCTGAGAAAT 420  
DB 361 TGCCTTAACACCTCTTGCGGATTCGATCCGTTTCCTTTATGCTAATTAATCTGAGAAAT 420

QY 421 GTTATTAACAGAACTCCTAATTTCTTACTGATGACAAACATCTGATTAATTAACCAAC 480  
DB 421 GTTATTAACAGAACTCCTAATTTCTTACTGATGACAAACATCTGATTAATTAACCAAC 480

QY 481 TGTGTACTGACTAATTTTCACTGCTTCCGGATGTTCTCTTCACCTGACCGGATGTA 540  
DB 481 TGTGTACTGACTAATTTTCACTGCTTCCGGATGTTCTCTTCACCTGACCGGATGTA 540

QY 541 TCCAGCTGTGATTT 555  
DB 541 TCCAGCTGTGATTT 555

RESULT 2  
AR560977 555 bp DNA linear PAT 08-OCT-2004  
LOCUS AR560977  
DEFINITION Sequence 5 from patent US 6756221.  
ACCESSION AR560977  
VERSION AR560977.1 GI:53973496  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 555)  
TITL Unclassified.  
AUTHORS Yamaguchi, S.  
JOURNAL Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof  
Patent: US 6756221-A 5 29-JUN-2004;  
Amano Enzyme Inc.; Aichi;  
JPX;

FEATURES  
Source Location/Qualifiers  
1..555 /organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN  
Query Match 100.0%; Score 555; DB 2; Length 555;  
Best Local Similarity 100.0%; Pred. No. 3.1e-167;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGAAAT 60

DB 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGAAAT 60

QY 61 CAGCTTTGGGATCCTCTACAGGCGCTCTCCACATGCATCATCTTCAATTCCTGTAGAC 120  
DB 61 CAGCTTTGGGATCCTCTACAGGCGCTCTCCACATGCATCATCTTCAATTCCTGTAGAC 120

QY 121 GGATGTATGCAAGAGCCCAATAGATGAGCAATCTTAATGAACAACGGCTATGACTGT 180  
DB 121 GGATGTATGCAAGAGCCCAATAGATGAGCAATCTTAATGAACAACGGCTATGACTGT 180

QY 181 GAAATCAATTTGTATACGGAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 240  
DB 181 GAAATCAATTTGTATACGGAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 240

QY 241 AGCTACCAAGTTCATATTTGTAGTATTAATTAATGCTCCGGAGTAAAGGAAAAAGA 300  
DB 241 AGCTACCAAGTTCATATTTGTAGTATTAATTAATGCTCCGGAGTAAAGGAAAAAGA 300

QY 301 ATTATGATCCTTCACATATTTTCAAGCGGCTCTGTAAACAGATACAGATGAGAAACGCT 360  
DB 301 ATTATGATCCTTCACATATTTTCAAGCGGCTCTGTAAACAGATACAGATGAGAAACGCT 360

QY 361 TGCCTTAACACCTCTTGCGGATTCGATCCGTTTCCTTTATGCTAATTAATCTGAGAAAT 420  
DB 361 TGCCTTAACACCTCTTGCGGATTCGATCCGTTTCCTTTATGCTAATTAATCTGAGAAAT 420

QY 421 GTTATTAACAGAACTCCTAATTTCTTACTGATGACAAACATCTGATTAATTAACCAAC 480  
DB 421 GTTATTAACAGAACTCCTAATTTCTTACTGATGACAAACATCTGATTAATTAACCAAC 480

QY 481 TGTGTACTGACTAATTTTCACTGCTTCCGGATGTTCTCTTCACCTGACCGGATGTA 540  
DB 481 TGTGTACTGACTAATTTTCACTGCTTCCGGATGTTCTCTTCACCTGACCGGATGTA 540

QY 541 TCCAGCTGTGATTT 555  
DB 541 TCCAGCTGTGATTT 555

RESULT 3  
AX113621 555 bp DNA linear PAT 30-MAY-2001  
LOCUS AX113621  
DEFINITION Sequence 5 from Patent EP1106696.  
ACCESSION AX113621  
VERSION AX113621.1 GI:13939799  
KEYWORDS  
SOURCE Chryseobacterium sp. No. 9670  
ORGANISM Chryseobacterium sp. No. 9670  
REFERENCE 1  
TITL Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof  
JOURNAL Patent: EP 1106696-A 5 13-JUN-2001;  
Amano Enzyme Inc. (JP)  
Flavobacteriales; Chryseobacterium.

FEATURES  
Source Location/Qualifiers  
1..555 /organism="Chryseobacterium sp. No. 9670"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:161487"

ORIGIN  
Query Match 100.0%; Score 555; DB 2; Length 555;  
Best Local Similarity 100.0%; Pred. No. 3.1e-167;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGAAAT 60  
DB 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGAAAT 60

QY 61 CAGCTTTGGGATCCTCTACAGGCGCTCTCCACATGCATCATCTTCAATTCCTGTAGAC 120

Db 61 CAGCTCTGGGAGTACCTCTAGCGGCTCTCAACCATGATCACTTCAGATATCTCTAGAC 120  
QY 121 GGATGTTATGCAAGAGCCCATAGATGAGCAAACTTATGAAACAGGCTATGACTGT 180  
Db 121 GGATGTTATGCAAGAGCCCATAGATGAGCAAACTTATGAAACAGGCTATGACTGT 180  
QY 181 GAAAAACAATTTGTATACGAAACCTAAAGCATCAACAGAACTTGCTGTGTGGCTGTG 240  
Db 181 GAAAAACAATTTGTATACGAAACCTAAAGCATCAACAGAACTTGCTGTGTGGCTGTG 240  
QY 241 AGCTACCAAGTGTGCAATTTGTAGCTATATAAAATGCTTCGGAGTAACGAAAAAGA 300  
Db 241 AGCTACCAAGTGTGCAATTTGTAGCTATATAAAATGCTTCGGAGTAACGAAAAAGA 300  
QY 301 ATTATGATGCTCTCACTATTTTCAAGGGGCTCTTAACAGATACAGATGGAGAAACGCT 360  
Db 301 ATTATGATGCTCTCACTATTTTCAAGGGGCTCTTAACAGATACAGATGGAGAAACGCT 360  
QY 361 TGGCTTAACAGCTCTTGGGATCTGCATCCGTTCCCTTATGCTAATCTGAGGAAAT 420  
Db 361 TGGCTTAACAGCTCTTGGGATCTGCATCCGTTCCCTTATGCTAATCTGAGGAAAT 420  
QY 421 GTTATTAACAGAACTCTAGTAATTTCTTACCTGTATGACAAACATTTGATCAATCAAC 480  
Db 421 GTTATTAACAGAACTCTAGTAATTTCTTACCTGTATGACAAACATTTGATCAATCAAC 480  
QY 481 TGTGTACTGACTAAATTTTCACTGCTTCGGAGTGTCTTCCTCACTGACCGGATGTA 540  
Db 481 TGTGTACTGACTAAATTTTCACTGCTTCGGAGTGTCTTCCTCACTGACCGGATGTA 540  
QY 541 TCCAGCTGTGATTT 555  
Db 541 TCCAGCTGTGATTT 555

RESULT 4  
BD015600 1080 bp DNA linear PAT 27-AUG-2002  
LOCUS Novel protein deamidase, microorganism producing the same, gene  
DEFINITION encoding the same, process for producing the same and use of the  
same.

ACCESSION BD015600  
VERSION BD015600.1 GI:22556737  
KEYWORDS JP 2001218590-A/4  
SOURCE Chryseobacterium sp. No. 9670  
ORGANISM Chryseobacterium sp. No. 9670  
Bacteria; Bacteroidetes; Flavobacteriales;  
Flavobacteriaceae; Chryseobacterium.

REFERENCE Yamaguchi, S.  
1 (bases 1 to 1080)  
AUTHORS Novel protein deamidase, microorganism producing the same, gene  
TITLE encoding the same, process for producing the same and use of the  
JOURNAL Patent: JP 2001218590-A 4 14-AUG-2001;  
AMANO ENZYME CO LTD  
COMMENT OS Chryseobacterium sp. No. 9670  
PN JP 2001218590-A/4  
PD 14-AUG-2001  
PF 04-DEC-2000 JP 2000368983  
PI SHOTARO YAMAGUCHI  
PC C12N15/09, A23J3/34, C07K1/13, C12N1/15, C12N1/19, C12N1/20, C12N1/  
PC 21, C12N5/10,  
PC C12N9/80//((C12N1/20, C12R1:01), (C12N9/80, C12R1:01), C12N15/00,  
PC C12N5/00  
CC Novel protein deamidase, microorganism producing the same, CC  
gene encoding  
the same, process for producing the same and use of the same  
FH Key (61), (1020)  
FT CDS Location/Qualifiers  
FT mat\_peptide (466), (1020).  
Location/Qualifiers

FEATURES  
source 1..1080  
/organism="Chryseobacterium sp. No. 9670"

ORIGIN  
/mol\_type="genomic DNA"  
/db\_xref="taxon:161487"

Query Match 100.0%; Score 555; DB 2; Length 1080;  
Best Local Similarity 100.0%; Pred. No. 3.4e-167;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGCAGTGTATATCTCGATGTAGTACATTAATTTCTTATTCATCAATTAAGAAAT 60  
Db 466 TTGGCAGTGTATATCTCGATGTAGTACATTAATTTCTTATTCATCAATTAAGAAAT 525  
QY 61 CAGCTCTGGGAGTCTTACGCGGCTCTGACATGATCAATTCAGATATCTGTAGAC 120  
Db 526 CAGCTCTGGGAGTCTTACGCGGCTCTGACATGATCAATTCAGATATCTGTAGAC 585  
QY 121 GGATGTTATGCAAGAGCCCATAGATGAGCAAACTTATGAAACAGGCTATGACTGT 180  
Db 586 GGATGTTATGCAAGAGCCCATAGATGAGCAAACTTATGAAACAGGCTATGACTGT 645  
QY 181 GAAAAACAATTTGTATACGAAACCTAAAGCATCAACAGAACTTGCTGTGTGGCTGTG 240  
Db 646 GAAAAACAATTTGTATACGAAACCTAAAGCATCAACAGAACTTGCTGTGTGGCTGTG 705  
QY 241 AGCTACCAAGTGTGCAATTTGTAGCTATATAAAATGCTTCGGAGTAACGAAAAAGA 300  
Db 706 AGCTACCAAGTGTGCAATTTGTAGCTATATAAAATGCTTCGGAGTAACGAAAAAGA 765  
QY 301 ATTATGATGCTCTCACTATTTTCAAGGGGCTCTTAACAGATACAGATGGAGAAACGCT 360  
Db 766 ATTATGATGCTCTCACTATTTTCAAGGGGCTCTTAACAGATACAGATGGAGAAACGCT 825  
QY 361 TGGCTTAACAGCTCTTGGGATCTGCATCCGTTCCCTTATGCTAATCTGAGGAAAT 420  
Db 826 TGGCTTAACAGCTCTTGGGATCTGCATCCGTTCCCTTATGCTAATCTGAGGAAAT 885  
QY 421 GTTATTAACAGAACTCTAGTAATTTCTTACCTGTATGACAAACATTTGATCAATCAAC 480  
Db 886 GTTATTAACAGAACTCTAGTAATTTCTTACCTGTATGACAAACATTTGATCAATCAAC 945  
QY 481 TGTGTACTGACTAAATTTTCACTGCTTCGGAGTGTCTTCCTCACTGACCGGATGTA 540  
Db 946 TGTGTACTGACTAAATTTTCACTGCTTCGGAGTGTCTTCCTCACTGACCGGATGTA 1005  
QY 541 TCCAGCTGTGATTT 555  
Db 1006 TCCAGCTGTGATTT 1020

RESULT 5  
AR560978 1080 bp DNA linear PAT 08-OCT-2004  
LOCUS AR560978  
DEFINITION Sequence 7 from patent US 6756221.  
ACCESSION AR560978  
VERSION AR560978.1 GI:53973497  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 1080)  
AUTHORS Yamaguchi, S.  
TITLE Protein-deamidating enzyme, microorganism producing the same, gene  
JOURNAL encoding the same, production process therefor, and use thereof  
Patent: US 6756221-A 7 29-OCT-2004;  
Amano Enzyme Inc.; Atch1;  
JPX;

FEATURES  
source 1..1080  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 555; DB 2; Length 1080;

Best Local Similarity 100.0%; Pred. No. 3.4e-167;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TTGGCAGGTGTAATTCCTGATGTAAGTACATTAATTCCTTATTCATCAATAAAGAAAT 60
Db 466 TTGGCAGGTGTAATTCCTGATGTAAGTACATTAATTCCTTATTCATCAATAAAGAAAT 525
QY 61 CAGTCTTGCGGTACCTCTACGGGCTCTCTACCATGCAATCATTGCAATATCCTGTAGAC 120
Db 526 CAGTCTTGCGGTACCTCTACGGGCTCTCTACCATGCAATCATTGCAATATCCTGTAGAC 585
QY 121 GGATGTTATGCAAGAGCCCATTAAGATGAGCAAAATCTTAATGAAACAACGGCTATGACTGT 180
Db 586 GGATGTTATGCAAGAGCCCATTAAGATGAGCAAAATCTTAATGAAACAACGGCTATGACTGT 645
QY 181 GAAAAACAATTTGTATACGAAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 240
Db 646 GAAAAACAATTTGTATACGAAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 705
QY 241 AGCTACACAGTTGCAATATGTTAGCTATTAATAATGCTTCCGAGTAAAGGAAAAAGA 300
Db 706 AGCTACACAGTTGCAATATGTTAGCTATTAATAATGCTTCCGAGTAAAGGAAAAAGA 765
QY 301 ATTATGATCCTTCACTATTTTCAAGCGGTCCTGTAAACAATACAGATGAGGAAACGCT 360
Db 766 ATTATGATCCTTCACTATTTTCAAGCGGTCCTGTAAACAATACAGATGAGGAAACGCT 825
QY 826 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCTTATGCTAATATCGCAGAAAT 885
Db 886 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCTTATGCTAATATCGCAGAAAT 945
QY 421 GTTATTAACAGAACTCTTAAGTATTTCTTACCTGTATACAAACATCTGATCAATACCAAC 480
Db 481 GTTATTAACAGAACTCTTAAGTATTTCTTACCTGTATACAAACATCTGATCAATACCAAC 945
QY 481 TGTGTACTGACTAATTTTCACTGCTTCCGATGTTCTCCTTCACTGACCGGATGTA 540
Db 946 TGTGTACTGACTAATTTTCACTGCTTCCGATGTTCTCCTTCACTGACCGGATGTA 1005
QY 541 TCCAGCTGTGGATTT 555
Db 1006 TCCAGCTGTGGATTT 1020
```

RESULT 6  
AX113623 1080 bp DNA linear PAT 30-MAY-2001  
LOCUS AX113623  
DEFINITION Sequence 7 from Patent EP1106696.  
ACCESSION AX113623  
VERSION AX113623.1 GI:13939800  
KEYWORDS  
SOURCE Chryseobacterium sp. No. 9670  
ORGANISM Bacteria; Bacteroidetes; Flavobacteriales; Flavobacteriaceae; Chryseobacterium.

REFERENCE  
AUTHORS Yamaguchi, S.  
TITLE Protein-deamidating enzyme, microorganism producing the same, gene encoding the same, production process therefor, and use thereof  
JOURNAL Patent: EP 1106696-A 7 13-JUN-2001;  
Amano Enzyme Inc. (JP)

FEATURES  
source  
1..1080  
/organism="Chryseobacterium sp. No. 9670"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:161487"  
61..1023  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37857.1"  
/db\_xref="GI:13939801"  
/translation="MKNLFLSMMAFVTVLTFNSCADNSNGNDEINGEKLSVNSDKLKD

mat\_peptide  
mat\_peptide  
/product="unnamed"

Query Match 100.0%; Score 555; DB 2; Length 1080;  
Best Local Similarity 100.0%; Pred. No. 3.4e-167;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TTGGCAGGTGTAATTCCTGATGTAAGTACATTAATTCCTTATTCATCAATAAAGAAAT 60
Db 466 TTGGCAGGTGTAATTCCTGATGTAAGTACATTAATTCCTTATTCATCAATAAAGAAAT 525
QY 61 CAGTCTTGCGGTACCTCTACGGGCTCTCTACCATGCAATCATTGCAATATCCTGTAGAC 120
Db 526 CAGTCTTGCGGTACCTCTACGGGCTCTCTACCATGCAATCATTGCAATATCCTGTAGAC 585
QY 121 GGATGTTATGCAAGAGCCCATTAAGATGAGCAAAATCTTAATGAAACAACGGCTATGACTGT 180
Db 586 GGATGTTATGCAAGAGCCCATTAAGATGAGCAAAATCTTAATGAAACAACGGCTATGACTGT 645
QY 181 GAAAAACAATTTGTATACGAAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 240
Db 646 GAAAAACAATTTGTATACGAAACCTTAAGGCATCAACAGAACTTGCTGTGGCGTGG 705
QY 241 AGCTACACAGTTGCAATATGTTAGCTATTAATAATGCTTCCGAGTAAAGGAAAAAGA 300
Db 706 AGCTACACAGTTGCAATATGTTAGCTATTAATAATGCTTCCGAGTAAAGGAAAAAGA 765
QY 301 ATTATGATCCTTCACTATTTTCAAGCGGTCCTGTAAACAATACAGATGAGGAAACGCT 360
Db 766 ATTATGATCCTTCACTATTTTCAAGCGGTCCTGTAAACAATACAGATGAGGAAACGCT 825
QY 826 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCTTATGCTAATATCGCAGAAAT 885
Db 886 TGCGTTAACACCTCTTGCGGATCTGCATCCGTTTCTTATGCTAATATCGCAGAAAT 945
QY 421 GTTATTAACAGAACTCTTAAGTATTTCTTACCTGTATACAAACATCTGATCAATACCAAC 480
Db 481 GTTATTAACAGAACTCTTAAGTATTTCTTACCTGTATACAAACATCTGATCAATACCAAC 945
QY 481 TGTGTACTGACTAATTTTCACTGCTTCCGATGTTCTCCTTCACTGACCGGATGTA 540
Db 946 TGTGTACTGACTAATTTTCACTGCTTCCGATGTTCTCCTTCACTGACCGGATGTA 1005
QY 541 TCCAGCTGTGGATTT 555
Db 1006 TCCAGCTGTGGATTT 1020
```

RESULT 7  
AB046594 1380 bp DNA linear BCT 27-APR-2001  
LOCUS AB046594  
DEFINITION Chryseobacterium proteolyticum prga gene for protein-glutaminase,  
complete cds.  
ACCESSION AB046594  
VERSION AB046594.1 GI:12597204  
KEYWORDS  
SOURCE Chryseobacterium proteolyticum  
ORGANISM Bacteria; Bacteroidetes; Flavobacteriales; Flavobacteriaceae; Chryseobacterium.

REFERENCE  
AUTHORS Yamaguchi, S., Jeenes, D.J. and Archer, D.B.  
TITLE Protein-glutaminase from Chryseobacterium proteolyticum, an enzyme that deamidates glutamyl residues in proteins. Purification, characterization and gene cloning  
JOURNAL Eur. J. Biochem. 268 (5), 1410-1421 (2001)  
PubMed 11231294





[illegible]

Db	427	TACGAAGTCCGTCAGGTTCACTATCTATGATPAACAATCATATGTAATACCAATTGTGTGA	486
QY	487	CTGACTAAATTTTTCACAGCTTTCGAGATGTTCTCTTCACTGACCCGATGTATCCAGC	546
Db	487	TTAAACATATTTCTCATCCCTTTTACAGATGTCTCTCCACAGACCAAGTGTAGCAAGC	546
QY	547	TGTGATTTT 555	
Db	547	TGTGATTTT 555	
RESULT 10			
ARS70507		555 bp	DNA
LOCUS			linear
DEFINITION	Sequence 5 from patent US 6770469.		PAT 14-DEC-2004
ACCESSION	ARS70507		
VERSION	ARS70507.1	GI:56571305	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 555)		
AUTHORS	Yamaguchi, S. and Matsunura, A.		
TITLE	Protein-deamidating enzyme, gene encoding the same, production process thereof, and use thereof;		
JOURNAL	Patent: US 6770469-A 5 03-AUG-2004;		
	Amano Pharmaceutical Co., Ltd.; Aichi;		
	JPX;		
FEATURES	location/qualifiers		
source	1..555		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	62.0%;	Score 344.2;	DB 2;
Best Local Similarity	76.7%;	Pred. No. 3.2e-99;	
Matches	421;	Conservative 0;	Mismatches 128;
		Indels 0;	Gaps 0;
QY	7	AGTGAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATTAAGATCAGTCT	66
Db	7	AGTGTATTCCTGATCTGGCAACGTGAAAGTTATTTACCCAGATCAAAAACAGGCT	66
QY	67	TGCGTACCTCTACGGCGTCTCTACCATGATCATCATTCAATATCTCTGTAGACGATGT	126
Db	67	TGCGGAATCTTACAGCATCTTCTCTGTATCACCTTCAGATATCCGGTTGACGAGTGT	126
QY	127	TATGGAAGGCCATPAAGATGAGACAAATCTTAATGAACAAGCGGTATGACTGTGAATAA	186
Db	127	TATGGAAGGGCTCACAAATAGACAAATCTTAATGAACCGCGGTATGACTGTGAATAA	186
QY	187	CAATTTGTATACGAAACCTAAAGGCATCAACAGAACTTCTGTGTGCGTGTGAGCTAC	246
Db	187	CAGTTCGTATATGTATATCTAGAGCTCTACAGGAACATCGTGTGTATCATGGGTATAT	246
QY	247	CAGCTTCAATTTGTGTAGCTATATAAATGCTTCCGAGTGAACGAAAAAAGATTATT	306
Db	247	CAGGTAGCAATTTGTGTAGCTTCAAAAATGCTTCCAGAAATGTTGAAAAAAGATCATTA	306
QY	307	GATCCTTCACTATTTTCAAGCGGTCCTGTAAACAGATACAGCATGGAGAAACGCTTGCTT	366
Db	307	GATCCTTCACTATTTTCAAGCGGTCCTGTAAACAGATTTGCAATGGAGAGCTGATGTACC	366
QY	367	AACACCTCTTGGGAGTGTGATCCGTTTCTTTATGCTAAATCTGACAGAAATGTTTAT	426
Db	367	AACACAGAGTGGGATCTGCGCTGTATCTTCTTACGACCAATACAGCAGAAATGTTTAC	426
QY	427	TACAGAACTCTAATTAATTTTCTCTGTATGACAAATCTGATCAATATACCACTGTGTGA	486
Db	427	TACAGAAAGTCCTGACGTTCACTATCTATGATTAACAATATGTGAATATACAAATGTGTGA	486
QY	487	CTGACTAAATTTTCTACTGCTTCCGAGATGTTCTCTTACCTGCACCGGATGTATCCAGC	546

Db 487 TTAACAATATTCATCCCTTCAGAGTGTCTTCCTCCAGACCAAGTGAACAAGC 546  
QY 547 TGTGATTT 555  
Db 547 TGTGATTT 555

RESULT 11  
LOCUS ARI59962 1080 bp DNA linear PAT 17-OCT-2001  
DEFINITION Sequence 11 from patent US 6251651.  
ACCESSION ARI59962  
VERSION ARI59962.1 GI:16222841  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1080)  
AUTHORS Yamaguchi, S. and Matsunura, A.  
TITLE Protein-deamidating enzyme, gene encoding the same, production process thereof, and use thereof  
JOURNAL Patent: US 6251651-A 11 26-JUN-2001;  
FEATURES  
source location/Qualifiers  
1. 1080  
/organism="unknown"  
/mol\_type="unassigned DNA"

## ORIGIN

Query Match 62.0%; Score 344.2; DB 2; Length 1080;  
Best Local Similarity 76.7%; Pred. No. 3.5e-99;  
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 7 AGGTATTCCTGATGCTAGCTACATTAATCTTTATTCATGAATAAAGATCAGTCT 66  
Db 469 AGGTATTCCTGATGCTAGCTAGCAAGCTGACAGATTATTTACCAAGATCAAAACCGGCT 528  
QY 67 TGGGTACCTCTAAGCGGCTCTCAGCATGATCAGATTCAGATTCCTGTAGACGATGT 126  
Db 529 TGGGAACTTCTAAGAGATCTTCTCCTGTATCACCCTTCAGATTCGGGTGACGATGT 588  
QY 127 TATGCAAGGCCATTAAGATGAGACAAATCTTAATGAACAAGCGCTATGACTGTGAAAA 186  
Db 589 TATGCAAGGGCTCACAAATGAGACAAATCTTAATGAACAAGCGCTATGACTGTGAAAA 648  
QY 187 CAATTTGTATAGGAAACCTAAAGCATCAACAGAACTTGCTGTGGCGGTGAGCTAC 246  
Db 649 CAATTTGTATAGTATATCTGAAGCTTCTTACAGAAACATGCTGTATCATGATATAT 708  
QY 247 CAGGTGCAATATTTGTAAGCTATTAATAATGCTTCGAGTAAACGAAAAAAGATTAT 306  
Db 709 CAGGTGCAATTTTGTGAAGCTTCAAAAATGCTTCAGGAATTTGTAAGAAAAAATCAT 768  
QY 307 GATCTTCACTATTTTCAAGCGGCTCTGAACAGATACAGATGAGAAACGCTTGCGTT 366  
Db 769 GATCTTCACTATTTTCAAGCGGCTCTGAACAGATACAGATGAGAAACGCTTGCGTT 828  
QY 367 AACACCTTGGCGGATCTGATCCGTTCTCTTAATGCTAATCTGCAAGAAATGTTAT 426  
Db 829 AACACAGCTGCGGATCTGCTGTATCTTCTTACGCCAATACAGAGAAATGTTTAC 888  
QY 427 TACAGAGTCTAGTATTTCTTACCTGTATGACAAACATCTGATCAATACCACTGTGA 486  
Db 889 TACAGAGTCTGACAGTTCATTAATCTGTATGATTAACAATGTATGATTAATGTGA 948  
QY 487 CTGACTAAATTTTCACTGCTTTCGAGATGTTCTCTTCACTGACCTGACCGAGTGTAC 546  
Db 949 TTAACAATATTCATCCCTTCAGAGTGTCTTCCTCCAGACCAAGTGAACAAGC 1008  
QY 547 TGTGATTT 555  
Db 1009 TGTGATTT 1017

RESULT 12  
LOCUS E59342 1080 bp DNA linear PAT 18-JUN-2001  
DEFINITION Novel protein deamidation enzyme, gene encoding it, process for producing the same, and utilization thereof.  
ACCESSION E59342  
VERSION E59342.1 GI:13023309  
KEYWORDS JP 2000050887-A/8.  
SOURCE Chryseobacterium gleum  
ORGANISM Chryseobacterium gleum

## REFERENCE

1 (bases 1 to 1080)  
AUTHORS Shotaro, Y. and Akira, M.  
TITLE Novel protein deamidation enzyme, gene encoding it, process for producing the same, and utilization thereof  
JOURNAL Patent: JP 2000050887-A 8 22-FEB-2000;  
FEATURES  
source location/Qualifiers  
1. 1080  
/organism="Chryseobacterium gleum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:250"

## ORIGIN

Query Match 62.0%; Score 344.2; DB 2; Length 1080;  
Best Local Similarity 76.7%; Pred. No. 3.5e-99;  
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 7 AGGTATTCCTGATGCTAGCTACATTAATCTTTATTCATGAATAAAGATCAGTCT 66  
Db 469 AGGTATTCCTGATGCTAGCTAGCAAGCTGACAGATTATTTACCAAGATCAAAACCGGCT 528  
QY 67 TGGGTACCTCTAAGCGGCTCTCAGCATGATCAGATTCAGATTCCTGTAGACGATGT 126  
Db 529 TGGGAACTTCTAAGAGATCTTCTCCTGTATCACCCTTCAGATTCGGGTGACGATGT 588  
QY 127 TATGCAAGGCCATTAAGATGAGACAAATCTTAATGAACAAGCGCTATGACTGTGAAAA 186  
Db 589 TATGCAAGGGCTCACAAATGAGACAAATCTTAATGAACAAGCGGCTATGACTGTGAAAA 648  
QY 187 CAATTTGTATAGGAAACCTAAAGCATCAACAGAACTTGCTGTGGCGGTGAGCTAC 246  
Db 649 CAATTTGTATAGTATATCTGAAGCTTCTTACAGAAACATGCTGTATCATGATATAT 708  
QY 247 CAGGTGCAATATTTGTAAGCTATTAATAATGCTTCGAGTAAACGAAAAAAGATTAT 306  
Db 709 CAGGTGCAATTTTGTGAAGCTTCAAAAATGCTTCAGGAATTTGTAAGAAAAAATCAT 768  
QY 307 GATCTTCACTATTTTCAAGCGGCTCTGAACAGATACAGATGAGAAACGCTTGCGTT 366  
Db 769 GATCTTCACTATTTTCAAGCGGCTCTGAACAGATACAGATGAGAAACGCTTGCGTT 828  
QY 367 AACACCTTGGCGGATCTGATCCGTTCTCTTAATGCTAATCTGCAAGAAATGTTAT 426  
Db 829 AACACAGCTGCGGATCTGCTGTATCTTCTTACGCCAATACAGAGAAATGTTTAC 888  
QY 427 TACAGAGTCTAGTATTTCTTACCTGTATGACAAACATCTGATCAATACCACTGTGA 486  
Db 889 TACAGAGTCTGACAGTTCATTAATCTGTATGATTAACAATGTATGATTAATGTGA 948



JOURNAL Submitted (28-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
REFERENCE 3 (bases 1 to 239553)  
AUTHORS Rat Genome Sequencing Consortium.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
COMMENT On May 10, 2003 this sequence version replaced gi:25091888. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## ----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: GHVY  
Center clone name: CH230-100L14

## ----- Summary Statistics

Assembly program: Atlas 3.0:  
Consensus quality: 223709 bases at least Q40  
Consensus quality: 226612 bases at least Q30  
Estimated insert size: 228176; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length.  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 230323: contig of 230323 bp in length  
\* 230324 230423: gap of unknown length  
\* 230424 231606: contig of 1183 bp in length  
\* 231607 231706: gap of unknown length  
\* 231707 233074: contig of 1368 bp in length  
\* 233075 233174: gap of unknown length  
\* 233175 239553: contig of 6379 bp in length.

## FEATURES

## source

1..239553  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-100L14"

## misc\_feature

1..1813  
/note="wgs contig"

## misc\_feature

11157..12737  
/note="wgs contig"

## gap

230324..230423  
/estimated\_length=unknown

## gap

231607..231706  
/estimated\_length=unknown

## gap

233075..233174  
/estimated\_length=unknown

## ORIGIN

Query Match 7.7%; Score 43; DB 12; Length 239553;  
Best Local Similarity 60.9%; Pred. No. 0.12;  
Matches 70; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 316 CTAATTTCAAGCGGTCCTGTAACAGATGAGAAAGCTTGCTTAACCTCT 375  
DB 61050 CTCGCTCACTCTGTCATTAACAAAGAACTTCATCACTCACTGAGAACTCAAT 61109  
QY 376 TCGGATTCGACGCCGTTCTTCTTATGCTAATGACGAGAAATGTTATTACA 430  
DB 61110 AATGCTCTGTACCGTTTCTGTTATGAGAAAGAACTGAGAAATGCAATTACA 61164

## RESULT 15

AF429315/c 125020 bp DNA linear PRI 18-JAN-2002  
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.  
ACCESSION AF429315  
VERSION AF429315.1 GI:17646244  
KEYWORDS

## SOURCE

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 125020)  
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,

Potter,N.T., Ross,C.A. and Marzolis,R.L.  
A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2

Nat. Genet. 29 (4), 377-378 (2001)  
11694876

2 (bases 1 to 125020)  
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.

Direct Submission  
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical

Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

## JOURNAL

## AUTHORS

## PUBMED

## REFERENCE

## TITLE

## JOURNAL

## FEATURES

## source

1..125020  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q24.3; between D16S520 and WI-12410"  
/note="Isolated from a patient with Huntington's  
Disease-Like 2 (HDL2)"  
complement(35581..35746)  
/rpt\_type=tandem  
/rpt\_unit\_seq="ctg"  
complement(<36507..36887)  
/gene="JPH3"  
/note="synonym: JPH3"  
complement(<36507..36887)  
/gene="JPH3"  
/product="Junctophilin 3"  
complement(<36507..36887)  
/gene="JPH3"  
/note="component of the junctional complex between plasma  
membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="Junctophilin 3"  
/protein\_id="AA140941.1"  
/db\_xref="GI:17646245"  
/translation="MSGGRRFVDDGGSYCGMEGKAKHGVCCTPKQGGYTSWS  
HGFVAGVTPWSGNTYGTAAQGRHGLSKKRWYKGMWTHGFKGRYVRECAQ  
NKAKEGWSNGLDQGYETYSDC"

## repeat\_region

## gene

## mRNA

## CDS

## ORIGIN

Query Match 7.4%; Score 41; DB 5; Length 125020;  
Best Local Similarity 12.8%; Pred. No. 0.5; Mismatches 154; Indels 0; Gaps 0;  
Matches 47; Conservative 154; Mismatches 165; Indels 0; Gaps 0;

```
QY      130  GCAAGGCCCATAGATGAGACAATCTTATGAAACAAGCTATGACTGTGAAAAACA 189
Db      66546  GSSWKAMTCTYAMWMMWMMWRMAARMAAGAGAGMGWRMRRTYWRYYRRRMRGRSRYYWA 66487
QY      190  TTGTATACGAAACCTAAAGCATCAACAGAACTTGCTGTGTGGCGGTGAGCTACAC 249
Db      66486  MYAGYGRRRGRGWYRCAYMWMWSRKATGKAMGAMYGWYTRWVKYKTKWTKKTKGYTCW 66427
QY      250  GTTGCAATATTTGTAAGCTATAAAAAATGCTTCGAGATAAGAAAAAGAAATTATTGAT 309
Db      66426  GYMMMKMAATWMTTWTYYAMRKYTKAATYKARWCMKCMWMMWMMKAKMSWMAWC 66367
QY      310  CCTTCACTATTTTCAAGCGGTCTGTAACAGATACAGATGAGAAACGCTTGCGTTAC 369
Db      66366  YGGMASMTWCASKSASKYRKGYSKYWMKSYSRGSWSTTYRARAARMMAARRRRWGR 66307
QY      370  ACCTCTGGCGATCTGCATCGTTCTCTTATGCTAATACTGCAGGAATGTTATTAC 429
Db      66306  ATSRSKKGMKWKAAGMRMTMAARAMCAGSSWRMSASRWAGSKRRWSYTSYWCYAC 66247
QY      430  AGAGTCTCTAGTAATTTCTTACCTGTATGACACACATCTGATCAATACCAACTGTACTG 489
Db      66246  AMAAAMAMWMAAAATYRGYCKSMRGAMRGAMRWMAAMWRRAKCMRGGYKWCYRGS 66187
QY      490  ACTAAA 495
Db      66186  ASRWWR 66181
```

Search completed: September 16, 2006, 15:58:07  
Job time : 3723 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd

OM nucleic - nucleic search, using sw model

Run on: September 16, 2006, 14:41:19 ; Search time 514 Seconds  
(without alignments)  
7528.401 Million cell updates/sec

```
Title: US-10-815-774-5
Perfect score: 555
Sequence: 1 ttggcgagtgtaattcctga.....atgtatccagctgtgatt 555
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5244920 seqs, 3486124231 residues  
Total number of hits satisfying chosen parameters: 10489840

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

1: geneseqn\_8:\*  
2: geneseqn\_1980s:\*  
3: geneseqn\_1990s:\*  
4: geneseqn\_2000s:\*  
5: geneseqn\_2001bs:\*  
6: geneseqn\_2002bs:\*  
7: geneseqn\_2002as:\*  
8: geneseqn\_2002bs:\*  
9: geneseqn\_2003bs:\*  
10: geneseqn\_2003cs:\*  
11: geneseqn\_2003ds:\*  
12: geneseqn\_2004bs:\*  
13: geneseqn\_2004bs:\*  
14: geneseqn\_2005s:\*  
15: geneseqn\_2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match.	Length	DB	ID	Description
1	555	100.0	555	4	AAE90280	AAE90280 Nucleotid
2	555	100.0	1080	4	AAE90281	AAE90281 Nucleotid
3	430.2	77.5	963	14	AAE01515	AAE01515 Corynefo
4	344.2	62.0	555	3	AAZ49444	AAZ49444 Chryseoba
5	344.2	62.0	1080	3	AAZ49445	AAZ49445 Chryseoba
6	41.6	7.5	2000	11	ACJ37108	ACJ37108 Rice stre
7	39.8	7.2	11000	12	ADQ24435_3	ADQ24435_3
8	39	7.0	1105	12	ADQ22883	ADQ22883 Human so
9	37	6.7	2000	8	ADAI1938	ADAI1938 Rice gene
C 9	37	6.7	15865	4	AAK05875	AAK05875 Human dig
C 10	37	6.5	464	14	ACLS8971	ACLS8971 Human co
11	36.2	6.5	547	6	ABN61671	ABN61671 Human can
12	36.2	6.5	19576	6	ABN170575	ABN170575 Chemical
C 13	36	6.5	19576	6	AAAG61258	AAAG61258 Human gen
C 14	35.8	6.5	65359	6	AAAD6665	AAAD6665 Human tra
C 15	35.8	6.4	2489	3	AAZ90461	AAZ90461 E. canis
C 16	35.6	6.4	2000	11	ACJ35887	ACJ35887 Rice stre
C 17	35.4	6.4	9300	5	ABAI5594	ABAI5594 Human ner
C 18	35.4	6.4	9300	5	ABAI5594	ABAI5594 Human ner

19	35.2	6.3	2000	11	AC135363
20	35	6.3	4383	4	ABL18719
21	35	6.3	4590	5	AAH24065
22	35	6.3	106416	4	ABL18718
23	34.8	6.3	1751	2	AAV40300
24	34.8	6.3	2000	8	ADA71938
25	34.8	6.3	2000	11	AC137108
26	34.8	6.3	50000	6	ABL56201
27	34.4	6.2	777	8	ACEF7327
28	34.4	6.2	1072	2	AAV74746
29	34.4	6.2	1570	5	ABAI8492
30	34.4	6.2	3510	13	ADT47537
31	34.4	6.2	3943	4	AA531317
32	34.4	6.2	3943	6	ABO65641
33	34.4	6.2	3943	10	ADCI10663
34	34.4	6.2	12582	4	AAK84357
35	34.4	6.2	23821	3	AAA92469
36	34.2	6.2	2019	4	AAFE2537
37	34.2	6.2	4197	2	AAO99430
38	34	6.1	511	13	ADQ57425
39	34	6.1	801	13	ADK64118
40	34	6.1	1653	13	ADT16715
41	34	6.1	2286	10	ADB63717
42	34	6.1	2669	5	ABV24889
43	34	6.1	2669	11	ACN91876
44	34	6.1	3212	10	ADC30112
45	34	6.1	3581	4	AAO9377
					AAO9377 Human ves

## ALIGNMENTS

CC	AAFP90280
CC	standard; DNA; 555 BP.
CC	AAFP90280
CC	AAF90280;
CC	06-AUG-2003 (revised)
CC	22-AUG-2001 (first entry)
CC	Nucleotide sequence of a protein-deamidating enzyme.
CC	Protein-deamidating enzyme; mineral absorption; food allergy; dough;
CC	bakery; confectionery; ss.
CC	Chryseobacterium sp.
CC	EP1106696-A1.
CC	13-JUN-2001.
CC	04-DEC-2000; 2000EP-00310768.
CC	03-DEC-1999; 99JP-00345044.
CC	(AMANO-) AMANO ENZYME INC.
CC	Yamaguchi S;
CC	WPI; 2001-376907/40.
CC	P-PADB; AAB84386.
CC	New enzyme for use in e.g. bakery has an ability to deaminate amido
CC	groups in a protein.
CC	Claim 8; Page 22; 43pp; English.
CC	The present sequence encodes a protein-deamidating enzyme from
CC	Cryoseobacterium sp. number 9670. The enzyme is able to deaminate amido
CC	groups in a protein by directly acting upon the amido groups without
CC	cutting peptide bonds and without cross-linking the protein. The enzyme
CC	thus reduces the mineral sensitivity of the protein and increases the

soluble mineral content in the protein-mineral solution, improving the absorption of minerals in the human body. The enzyme reduces or removes the toxicity and allergenic property of the protein in a food e.g. allergy. The enzyme is useful for the improvement of dough in the field of bakery and confectionery e.g. for the production of crackers, biscuits, cookies, pizza pies or crusts of pie; in producing soybean protein products, in various food articles e.g. meat or fish products and noodles; and for improving functionality of plant or animal protein.

(Updated on 06-AUG-2003 to correct OS field.)

Sequence 555 BP; 166 A; 119 C; 107 G; 163 T; 0 U; 0 Other;

Query Match 100.0%; Score 555; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. No. 1e-159;

Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TTGGCGAGTGAATTCCTGATGATGATCACTTAATCTTTATTCATCAATTAAGAAAT 60

1 TTGGCGAGTGAATTCCTGATGATGATCACTTAATCTTTATTCATCAATTAAGAAAT 60

61 CAGTCTTGGGATCCTTCAAGCGGCTCTCAACATGATCAATTCAGATATCCTGTAGAC 120

61 CAGTCTTGGGATCCTTCAAGCGGCTCTCAACATGATCAATTCAGATATCCTGTAGAC 120

121 GGATGTTATGCAAGAGCCCTAAGATGAGCAAAATCTTAAGCAACGGCTATGACTGT 180

121 GGATGTTATGCAAGAGCCCTAAGATGAGCAAAATCTTAAGCAACGGCTATGACTGT 180

181 GAAAAACAATTTGATACGAAACCTAAAGCATCAACAGAACTTGTGTGCGCTGG 240

181 GAAAAACAATTTGATACGAAACCTAAAGCATCAACAGAACTTGTGTGCGCTGG 240

241 AGTACCAAGTTCATATTTGTTAAGTATTAATAATCTTCCGAGTAACGAAAAAGA 300

241 AGTACCAAGTTCATATTTGTTAAGTATTAATAATCTTCCGAGTAACGAAAAAGA 300

301 ATTATGATCTTCACTATTTTCAAGCGGCTCTCAACATGATGAGAAAGCT 360

301 ATTATGATCTTCACTATTTTCAAGCGGCTCTCAACATGATGAGAAAGCT 360

361 TTGGTAAACACTTCTGCGATGATGATCGGTTCTCTTATGATTAATGACAGAAAT 420

361 TTGGTAAACACTTCTGCGATGATGATCGGTTCTCTTATGATTAATGACAGAAAT 420

421 GTTATTAACAAGAGTCTGATTAATTTTACCTGTATGACAACTTGATCAATACCAAC 480

421 GTTATTAACAAGAGTCTGATTAATTTTACCTGTATGACAACTTGATCAATACCAAC 480

481 TTGCTACTGATTAATTTTCACTGCTTCCGAGATGTTCTCTTCAACCGGAGATGTA 540

481 TTGCTACTGATTAATTTTCACTGCTTCCGAGATGTTCTCTTCAACCGGAGATGTA 540

541 TCCAGCTGTGATTT 555

541 TCCAGCTGTGATTT 555

AAAF90281 standard; DNA; 1080 BP.

AAAF90281;

06-AUG-2003 (revised)

22-AUG-2001 (first entry)

Nucleotide sequence of a protein-deamidating enzyme.

Protein-deamidating enzyme; mineral absorption; food allergy; dough;

bakery; confectionery; ss.

Chryseobacterium sp.

Key Location/Qualifiers

CDS 61..1023

FT /tag= a

FT /product= "protein-deamidating enzyme"

EP110696-A1.

13-JUN-2001.

04-DEC-2000; 2000EP-00310768.

03-DEC-1999; 99UP-00345044.

(AMANO-) AMANO ENZYME INC.

Yamaguchi S;

WPI; 2001-376907/40.

P-PSDB; AAB84387.

New enzyme for use in e.g. bakery has an ability to deaminate amido groups in a protein.

Example 11; Page 23; 43p; English.

The present sequence encodes a protein-deamidating enzyme from *Chryseobacterium* sp. number 9670. The enzyme is able to deaminate amido groups in a protein by directly acting upon the amido groups without cutting peptide bonds and without cross-linking the protein. The enzyme thus reduces the mineral sensitivity of the protein and increases the soluble mineral content in the protein-mineral solution, improving the absorption of minerals in the human body. The enzyme reduces or removes the toxicity and allergenic property of the protein in a food e.g. allergy. The enzyme is useful for the improvement of dough in the field of bakery and confectionery e.g. for the production of crackers, biscuits, cookies, pizza pies or crusts of pie; in producing soybean protein products, in various food articles e.g. meat or fish products and noodles; and for improving functionality of plant or animal protein.

(Updated on 06-AUG-2003 to correct OS field.)

Sequence 1080 BP; 371 A; 204 C; 203 G; 302 T; 0 U; 0 Other;

Query Match 100.0%; Score 555; DB 4; Length 1080;

Best Local Similarity 100.0%; Pred. No. 1.4e-159;

Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TTGGCGAGTGAATTCCTGATGATGATCACTTAATCTTTATTCATCAATTAAGAAAT 60

466 TTGGCGAGTGAATTCCTGATGATGATCACTTAATCTTTATTCATCAATTAAGAAAT 525

61 CAGTCTTGGGATCCTTCAAGCGGCTCTCAACATGATCAATTCAGATATCCTGTAGAC 120

526 CAGTCTTGGGATCCTTCAAGCGGCTCTCAACATGATCAATTCAGATATCCTGTAGAC 585

121 GGATGTTATGCAAGAGCCCTAAGATGAGCAAAATCTTAATGACAAACGGCTATGACTGT 180

586 GGATGTTATGCAAGAGCCCTAAGATGAGCAAAATCTTAATGACAAACGGCTATGACTGT 645

181 GAAAAACAATTTGATACGAAACCTAAAGCATCAACAGAACTTGTGTGCGCTGG 240

646 GAAAAACAATTTGATACGAAACCTAAAGCATCAACAGAACTTGTGTGCGCTGG 705

241 AGTACCAAGTTCATATTTGTTAAGTATTAATAATCTTCCGAGTAACGAAAAAGA 300

706 AGTACCAAGTTCATATTTGTTAAGTATTAATAATCTTCCGAGTAACGAAAAAGA 765

301 ATTATGATCTTCACTATTTTCAAGCGGCTCTGTAACAGATACAGCATGAGAAAGCT 360

766 ATTATGATCTTCACTATTTTCAAGCGGCTCTGTAACAGATACAGCATGAGAAAGCT 825

361 TTGGTAAACACTTCTGCGATGATGATCGGTTCTCTTATGATTAATGACAGAAAT 420

826 TTGGTAAACACTTCTGCGATGATGATCGGTTCTCTTATGATTAATGACAGAAAT 885



QY 421 GTTATTACGAAAGTCTAGTAATTTTACCTGTATGACAAATCTGATCAATACCAC 480  
DB 886 GTTATTACGAAAGTCTAGTAATTTTACCTGTATGACAAATCTGATCAATACCAC 945  
QY 481 TGTGTACTGACTAATTTTACCTGTATGACAAATCTGATCAATACCAC 540  
DB 946 TGTGTACTGACTAATTTTACCTGTATGACAAATCTGATCAATACCAC 1005  
QY 541 TCCAGCTGTGATTT 555  
DB 1006 TCCAGCTGTGATTT 1020

RESULT 3  
ID ABE01515 standard; DNA; 963 BP.  
XX ABE01515;  
AC ABE01515;  
XX 26-JAN-2006 (first entry)  
XX  
XX  
DE Coryneform protein production system-related DNA, SEQ ID NO:3.  
XX  
XX  
KM protein production; cell culture; genetically engineered microorganism;  
KW ds; gene.  
XX  
XX  
OS Unidentified.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..963  
FT /tag= a  
FT /product= "Undefined protein"  
PN MO2005103278-A1.  
XX  
PD 03-NOV-2005.  
PF 20-APR-2005; 2005MO-JP007518.  
XX  
XX  
-PR 20-APR-2004; 2004JP-00124196.  
PR 13-JUN-2005; 2005JP-00005896.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Date M, Kikuchi Y, Itaya H, Nakamura N;  
XX  
XX WPI, 2005-748042/76.  
DR P-PSDB; ABE01516.  
XX  
XX  
PT Producing heterologous protein by culturing Coryneform bacterium having  
PT expression gene construct comprising promoter sequence, nucleic acid  
PT encoding Tat system-dependent signal peptide region and nucleic acid  
PT encoding foreign protein.  
XX  
PS Disclosure; SEQ ID NO 3; 47pp; Japanese.  
XX  
XX  
CC The new invention relates to a method of producing heterologous protein,  
CC by culturing a Coryneform bacterium having an expression gene construct  
CC carrying, in the direction from the 5'-terminal to 3'-terminal, a  
CC promoter sequence functioning in the bacterium, a nucleic acid sequence  
CC encoding a Tat system-dependent signal peptide region and a nucleic acid  
CC sequence encoding a foreign protein, and allowing the bacterium to  
CC produce and secrete the protein. The signal peptide of (M1) comprises SEQ  
CC ID NO: 28-32. The signal peptide is a signal peptide of isomaltose dextran,  
CC which comprises SEQ ID NO: 6, or of trimethylamine N-oxidoreductase,  
CC which comprises SEQ ID NO: 8. The bacteria comprises gene that encodes  
CC Tat system secretion peptide such as tatA, tatB, tatC or tatZ. (M1) is  
CC useful for producing heterologous protein e.g. glutaminase or  
CC isomaltodextran. (M1) enables efficient production of heterologous  
CC protein. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences. The present

CC sequence is a DNA related to the invention.  
XX  
XX  
SQ Sequence 963 BP; 240 A; 290 C; 207 G; 226 T; 0 U; 0 Other;  
Query Match 77.5%; Score 430.2; DB 14; Length 963;  
Best Local Similarity 85.9%; Pred. No. 2.8e-121;  
Matches 477; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1 TTGGGAGTGTAAATTCCTGATGTAGTACATTAATTTCTTATTCATCAATTAAGAAAT 60  
DB 406 TTGGGAGTGTAAATTCCTGATGTAGTACATTAATTTCTTATTCATCAATTAAGAAAT 465  
QY 61 CAGTCTTGGGTAACCTTACGCGCTTCCATGATGATCAATTCATGATTCCTGTAGAC 120  
DB 466 CAGTCTTGGGTAACCTTACGCGCTTCCATGATGATCAATTCATGATTCCTGTAGAC 525  
QY 121 GGATGTTATGCAAGAGCCCATTAAGATGAGCAATTTTAAATGAACAGCGCTATGACTT 180  
DB 526 GGATGTTATGCAAGAGCCCATTAAGATGAGCAATTTTAAATGAACAGCGCTATGACTT 585  
QY 181 GAAAAACATTTGTATACGGAAGAACTTAAAGCATCAACAGAACTTGTGTGGCGTGG 240  
DB 586 GAAAAACATTTGTATACGGAAGAACTTAAAGCATCAACAGAACTTGTGTGGCGTGG 645  
QY 241 AGCTACCAAGTGTGCAATATTTGTTAACTTAAATATGCTTCCGAGTAAACGAAAAAGA 300  
DB 646 AGCTACCAAGTGTGCAATATTTGTTAACTTAAATATGCTTCCGAGTAAACGAAAAAGA 705  
QY 301 ATTATTTGATCTTCTACATTTTCAAGCGGCTCTGTAAACAGATACAGATGAGAAACGT 360  
DB 706 ATTATTTGATCTTCTACATTTTCAAGCGGCTCTGTAAACAGATACAGATGAGAAACGT 765  
QY 361 TSCGTTAACACCTTGGGAGATCTGATCCGCTTCTGTTATGCTAATCTGAGGAAAT 420  
DB 766 TSCGTTAACACCTTGGGAGATCTGATCCGCTTCTGTTATGCTAATCTGAGGAAAT 825  
QY 421 GTTATTACGAAAGTCTAGTAATTTTACCTGTATGACAAATCTGATCAATACCAC 480  
DB 826 GTTATTACGAAAGTCTAGTAATTTTACCTGTATGACAAATCTGATCAATACCAC 885  
QY 481 TGTGTACTGACTAATTTTACCTGTATGACAAATCTGATCAATACCAC 540  
DB 886 TGTGTACTGACTAATTTTACCTGTATGACAAATCTGATCAATACCAC 945  
QY 541 TCCAGCTGTGATTT 555  
DB 946 TCCAGCTGTGATTT 960

RESULT 4  
ID AA249494 standard; DNA; 555 BP.  
XX AA249494;  
AC AA249494;  
XX  
XX  
DT 15-SEP-2003 (revised)  
DT 04-APR-2000 (first entry)  
XX  
XX  
DE Chryseobacterium gleum protein-deamidating enzyme encoding DNA.  
XX  
XX  
KW Protein deamidating enzyme; soil bacterium; deamidation activity;  
KW amido group; carboxyl group; ammonia; cross linking; peptide bond;  
KW protein engineering; surface hydrophobicity; toxicity; allergic;  
KW mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.  
XX  
XX Chryseobacterium gleum; 'JCM 2410'.  
OS  
XX  
XX  
FH Key Location/Qualifiers  
FT 1..555  
FT /tag= a  
FT /product= "Protein deamidating enzyme"  
FT /function= "Deamidate amido groups in a protein"  
FT /note= "Improves protein function"

```

XX  EP976829-A2.
XX  02-FEB-2000.
XX
XX  04-JUN-1999; 99EP-00304367.
XX  PF
XX  04-JUN-1998; 98JP-00173940.
XX  PR
XX  (AMANO ) AMANO PHARM KK.
XX  PA
XX  Yamaguchi S, Matsuura A;
XX  WPI; 2000-118552/11.
XX  DR
XX  P-PSDB; AAY44582.
XX
XX  New enzyme for modifying and improving the function of proteins and/or
XX  peptides has deamidating activity without causing cross linking.
XX
XX  Claim 8; Page 23; 57pp; English.
XX
XX  The present sequence is the DNA encoding the protein-deamidating enzyme,
XX  isolated from a new strain of soil bacterium, Chryseobacterium gleum JCM
XX  2410. The enzyme exerts the deamidation activity by directly acting upon
XX  side chain amido groups in the protein in bonded state and releasing side
XX  chain carboxyl groups and ammonia. It can deaminate high molecular weight
XX  proteins, without cross linking and cleavage of peptide bonds, to improve
XX  protein function. This sequence is used for protein engineering, to cause
XX  an increase in surface hydrophobicity and improve the function of a plant
XX  or animal protein. It can also be used to remove or reduce toxicity of
XX  allergic property of proteins in food, decrease mineral sensitivity of
XX  protein, to allow greater absorption into the body and to solubilise
XX  calcium for use in drinks and mineral enhancing agents. (Updated on 15-
XX  SEP-2003 to standardise OS field)
XX
XX  Sequence 555 BP; 159 A; 122 C; 112 G; 162 T; 0 U; 0 Other;
XX
XX  Query Match 62.0%; Score 344.2; DB 3; Length 555;
XX  Best Local Similarity 76.7%; Pred. No. 5.6e-95;
XX  Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```

```

QY  487 CTGACTAAATTTTCAGCTGCTTCGAGATGTTCTCTTCACCGACCGAGATGATCCAGC 546
DB  487 TTTAAACATATTCATCCCTTTTCAGAGATGTTCTCTTCACCGACCGAGATGATCCAGC 546
QY  547 TGTGGATTT 555
DB  547 TGTGGATTT 555

RESULT 5
AAZ49495
AAZ49495 standard; DNA; 1080 BP.
XX
XX  AAZ49495;
AC  15-SEP-2003 (revised)
XX  DT
XX  04-APR-2000 (first entry)
XX  DT
XX  Chryseobacterium gleum protein-deamidating prepro-enzyme encoding gene.
XX  DE
XX  Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;
XX  KW amido group; carboxyl group; ammonia; cross linking; peptide bond;
XX  KW protein engineering; surface hydrophobicity; toxicity; allergic;
XX  KW mineral sensitivity; calcium; absorption; mineral enhancing agent; ds.
XX  OS
XX  Chryseobacterium gleum; 'JCM 2410'.
XX
XX  Key Location/Qualifiers
XX  FH CDS
XX  FT 61..1020
XX  FT /*tag= a
XX  FT /product= "C. gleum protein deamidating prepro-enzyme"
XX  FT /function= "Deaminate amido groups in a protein"
XX  FT 61..123
XX  FT /*tag= b
XX  FT /note= "Corresponds to pre region of the protein"
XX  FT mat_peptide
XX  FT 463..1017
XX  FT /*tag= c
XX  FT /label= "Mature protein deamidating enzyme"
XX  FT /note= "Improves protein function"
XX
XX  EP976829-A2.
XX
XX  02-FEB-2000.
XX  PD
XX
XX  04-JUN-1999; 99EP-00304367.
XX  PF
XX  04-JUN-1998; 98JP-00173940.
XX  PR
XX  (AMANO ) AMANO PHARM KK.
XX  PA
XX  Yamaguchi S, Matsuura A;
XX  PI
XX  WPI; 2000-118552/11.
XX  DR
XX  P-PSDB; AAY44583.
XX
XX  New enzyme for modifying and improving the function of proteins and/or
XX  peptides has deamidating activity without causing cross linking.
XX
XX  Example 26; Page 25; 57pp; English.
XX
XX  The present sequence is the gene encoding the protein-deamidating prepro-
XX  enzyme, isolated from a new strain of soil bacterium, Chryseobacterium
XX  gleum JCM 2410. The enzyme exerts the deamidation activity by directly
XX  acting upon side chain amido groups in the protein in bonded state and
XX  releasing side chain carboxyl groups and ammonia. It can deaminate high
XX  molecular weight proteins, without cross linking and cleavage of peptide
XX  bonds, to improve protein function. This sequence is used for protein
XX  engineering, to cause an increase in surface hydrophobicity and improve
XX  the function of a plant or animal protein. It can also be used to remove
XX  or reduce toxicity of allergic property of proteins in food, decrease
XX  mineral sensitivity of protein, to allow greater absorption into the body
XX  and to solubilise calcium for use in drinks and mineral enhancing agents.
XX  (Updated on 15-SEP-2003 to standardise OS field)

```



WP AD034435\_4 400001 510000  
 XX AD034435\_5 500001 610000  
 WP AD034435\_6 600001 634886

Query Match 7.2%; Score 39.8; DB 12; Length 110000;  
 Best Local Similarity 49.3%; Pred. No. 1.7;  
 Matches 104; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 18 TGATGAGCTACATTAATCTTATTCATCAATAAAGATCAGTCTTGCGGTACCTC 77  
 DB 85206 TGGTTGTTTAACTAACTAATCTTATTCACAGTTTCATATGACATTTACTTGTAGGCGTC 85265  
 QY 78 TACGGCGTCTCCATCCATGATCATCATTCAGATATCTTGAGCGGATTTATCCAGAGC 137  
 DB 85266 TTTTGTGCCAAAGCGATCGGTTCCATACAGCAGCCTCTGAGACATGTGTAGCTGATGAGC 85325  
 QY 138 CCATTAAGATGAGCAAAATCTTAAATGACCAACGGTATAGCTGTGAAAACAAATTTGTATA 197  
 DB 85326 AGATTAATAATATGGCTAGTCCAAATTAAGTATGCTGTAGATTAATAATGCAAAAGTGATT 85385  
 QY 198 CGGAACCTTAAGGATCAACAGGAACTTGC 228  
 DB 85386 TCCTAAACAGTATTCAAAAAAGAAATGC 85416

RESULT 8  
 ADQ22883  
 ID ADQ22883 standard; DNA; 1105 BP.  
 AC ADQ22883;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5703.  
 XX  
 KM soft tissue sarcoma; cyrostatic; gene therapy; vaccine; screening; human;  
 KM ds.  
 OS Homo sapiens.  
 OS  
 PN WO2004048938-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 26-NOV-2003; 2003WO-US038193.  
 XX  
 PR 26-NOV-2002; 2002US-0429739P.  
 XX  
 PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX  
 PI Aziz N, Ginsburg WM, Zlotnik A;  
 XX  
 DR WPI; 2004-441208/41.  
 XX  
 PT Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.  
 XX  
 PS Example 2; SEQ ID NO 5703; 210bp; English.  
 XX  
 CC The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cyrostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC DNA of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.  
 XX Sequence 1105 BP; 355 A; 204 C; 212 G; 334 T; 0 U; 0 Other;  
 SQ

Query Match 7.0%; Score 39; DB 12; Length 1105;  
 Best Local Similarity 52.8%; Pred. No. 0.41;  
 Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 182 AAAAACAATTTGTATACGGAACCTTAAGGATCAACAGGAACCTTGCTGTGCGGTGCA 241  
 DB 91 AAAAGATGTTTATAGAAATCTGGTTACTTCTTGAATATGAACTTTATGATTTAA 150  
 QY 242 GCTACACGTTGCAATATTTGTTAGCTATATAAAATGCTCCGAGTAAAGGAAAAA 301  
 DB 151 GCTACTCTGTTATATATGCGGTAGCTAGAGTAAAGCCAAAGAAAAAGGTTAAATTG 210  
 QY 302 TTATGATCCTCTCACTATTTTCAAGCGGTCTCTGTAAACG 340  
 DB 211 CTAAAGAAATTTCAACMAATTTAAGTTGCTTTTAAAGA 249

RESULT 9  
 ADA71938/C  
 ID ADA71938 standard; DNA; 2000 BP.  
 AC ADA71938;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Rice gene, SEQ ID 5263.  
 XX  
 KM Plant; bacterial infection; fungal infection; viral infection; rice;  
 KM gene; ds.  
 OS Oryza sativa.  
 OS  
 PN WO2003000898-A1.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;  
 XX  
 DR WPI; 2003-175290/17.  
 XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX  
 PS Claim 27; SEQ ID NO 5263; 899bp; English.  
 XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX  
 SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 6.7%; Score 37; DB 8; Length 2000;  
 Best Local Similarity 8.2%; Pred. No. 2.2;

```
Matches 43; Conservative 244; Mismatches 234; Indels 2; Gaps 1;
QY 33 AAATCTTTATTCATCAATAAAGATCAGTCTGGCGTACCTCAACGGCGCTCACC 92
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 894 RRRTYKSMWRMTWTMTKAMTMTTCMAKMTMTATGMAATMTMTTMTYTTCAMTCAK 835
QY 93 ATGCATCAATTCAGATATTCCTGTAGACGAGATTTATGCAAGACCCATAAGATGACAC 152
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 834 KYKMAATMTWTTCACMRATSMWRBAMAGMRKRYKMTKRAYMRWRMRCTWAGMARMKSR 775
QY 153 AATCTTAATGACAGACGGCTATGACTGTGAAAAACAATTTGTATACGGAACCTAAAGC 212
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 774 YRKMKKYATRYKMTMTMTWMSWRKWSYRWSGMRWRMSAMRYCSMKCAKTKYAS 715
QY 213 ATCAACAGAACTTGCTGTGTGGCTGAGACTCAACAGTTGCAATATTTGTAGACTATA 272
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 SARWTKRARSYRRRRMYKRGWTRRYRYMWSCTMTARMKRRKMAASMKSCMTWY 655
QY 273 AAATGCTTCGGAATACGAAAAAAGAAATTAATTCCTTCACTATTTTCAACGGATCC 332
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 654 RGARSMWWSKYSGSACCKKTRVTWTSYMTGYMWSYKSMSTSKSMYMGKMTCTWY 595
QY 333 TGTAACAGATACGACATGAGAAACGCTTCGTTAACACTCTTGCGGATCTGATCCGT 352
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 TS--MGSTRRSKMGSRMSRMYRMRKMKRKYMRKCTWRRCMCTRMGYMTY 537
QY 393 TTGCTCTTATGCTAATACGTGACAGAAATGTTTATTAACAAGTCTAGTAATCTTAACG 452
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 TSSRSMYTRRYKARYSKRRYMYKRYKTCMYTYGMYKSGYMYRGGYCKACCKCTTA 477
QY 453 GTATGACAACAATCTGATCAATACCAACTGTGACTGACTAAATTTTCACTGCTTCGG 512
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 MCKKAAYSGMMWYMYRKYSKWMSMTKYMWSMYKKCRSMKYAKGCGYKMTYSGY 417
QY 513 ATGTTCTCTTCACTGACCGGATGATTCACCTGTGATTT 555
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 MKWYTYGYSYKSRCYKMYMMYKGMWYMYYSAYSSMTWY 374

RESULT 10
AAK90825/c
ID AAK90825 standard; DNA; 15865 BP.
XX
AC AAK90825;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 4401.
XX
KM Human; digestive system antigen; gene therapy; cancer; appendicitis;
KM ulcerative colitis; infection; Hirschprung's disease; chronic colitis;
KM digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN MO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001324.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220693P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225577P.
PR 14-AUG-2000; 2000US-0225578P.
PR 14-AUG-2000; 2000US-022558P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239340P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
```

PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0244674P.
PR	08-NOV-2000;	2000US-0244675P.
PR	08-NOV-2000;	2000US-0244676P.
PR	08-NOV-2000;	2000US-0244677P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249266P.
PR	17-NOV-2000;	2000US-0249267P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	01-DEC-2000;	2000US-0250300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	08-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-502630/55.	
XX		
PT	Polynucleotides encoding digestive system antigens, useful for	
PT	diagnosing, treating, preventing and/or prognosing disorders of the	
XX	digestive system, particularly cancer and cancer metastases.	
XX		
PS	Disclosure; SEQ ID NO 4401; 9866p; English.	
XX		
CC	The present invention provides the protein and coding sequences of a	
CC	number of human digestive system antigens. These can be used in the	
CC	diagnosis, treatment and prevention of digestive system disorders,	
CC	including cancer, Meckel's diverticulum, bacterial or parasitic	
CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or	
CC	ulcerative colitis. The present sequence is a genomic DNA fragment	
CC	encoding a digestive system antigen of the invention	
XX		
SQ	Sequence 15865 BP; 4775 A; 3204 C; 3309 G; 4577 T; 0 U; 0 Other;	

	Query Match	6.7%;	Score 37;	DB 4;	Length 15665;
	Best Local Similarity	50.9%;	Pred. No.5.3;		
	Matches	88;	Conservative	0;	Mismatches 85; Indels 0; Gaps 0;
Oy	54	AAAGATCAAGTCTTTGGGATACCGTCACGGCAGTCCTCACCATGCATCACTTGAGATTGCC	113		
Dd	13604	AATGACTAGTCTCTCAAGAGTACTCTTGTACTTGAAAAATCTCATTTGAGAAAAAAAAA	13545		
Oy	114	TGTAGAAGATGTTATGCAAGAGCCCATTAAGATGAACAATCTTAAAGAACACGCGCTA	173		
Dd	13544	ACAAAAAACTTTATCATGGAGGAATAATTAATTGAGAAAATCATTAATGCCCTAACAGCTG	13485		
Oy	174	TGACTGTGAANAACAAATTTGTATACGGAACCCTAAAGGATCAACAGCAACTT	226		
Dd	13484	GTAACAGAGAAAAATATGTGTAATAATTAAGATATAGTAAATGTATCAGACTT	13452		
<b>RESULT_11</b>					
ID	ACL58971	standard; cDNA; 464 BP.			
XX	ACL58971;				
DT	24-MAR-2005	(first entry)			
DE	Human colon cancer differentially expressed polynucleotide, SEQ ID:5106.				
KW	Differential expression; diagnosis; therapy; drug screening; cancer;				
KM	neoplasm; colon tumor; breast tumor; pancreas tumor; cytostatic; vaccine;				
KX	ss.				
OS	Homo sapiens.				
PN	WO200500087-A2.				
PD	06-JAN-2005.				
PE	13-MAY-2004; 2004WO-USO15421.				
PR	03-JUN-2003; 2003JUS-0475872P.				
PA	(CHIR ) CHIRON CORP.				
PI	Randazzo F, Moier E, Escobedo J, Garcia PD;				
DR	WPI; 2005-075421/08.				
PT	New isolated polynucleotides, which are differentially expressed in colon				
PT	cancer cell, useful for treating cancer, e.g. colon cancer, breast				
PT	cancer, or pancreatic cancer.				
PS	Claim 1; SEQ ID NO 5106; 97bp; English.				
XX	The invention relates to 9672 polynucleotides (ACL53866-ACL63537) which				
CC	are differentially expressed in colon cancer cells. The invention also				
CC	relates to vectors and host cells comprising a differentially expressed				
CC	polynucleotide of the invention; a method for detecting a cancerous cell				
CC	by detection of a gene product of the polynucleotide; a method for				
CC	inhibiting a cancerous phenotype of a cell by inhibiting a gene product				
CC	of the polynucleotides; a method of treating an individual with cancer by				
CC	administration of a modulator of a gene product of the polynucleotides;				
CC	and an isolated antibody that specifically binds to a polypeptide encoded				
CC	by one of the 9672 polynucleotides. The polynucleotides, polypeptides,				
CC	antibodies, and methods are useful for the detection of cancerous cells;				
CC	for the diagnosis, prognosis and management of cancer; for the				
CC	identification of agents that modulate the phenotype of cancerous cells;				
CC	for the identification of therapeutic targets for cancer chemotherapy;				
CC	and for the treatment of cancer, especially colon cancer and metastasized				
CC	colon cancer, but also breast or pancreatic cancer. The polynucleotides				
CC	are also useful as a source of probes or primers for use in diagnostic				
CC	methods. The differentially expressed polynucleotides or their encoded				
CC	proteins can additionally be used as vaccines to modulate primary immune				







```
FT      /tag= k
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(19891, -)
FT      /tag= l
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(20272, A)
FT      /tag= m
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(20412, A)
FT      /tag= n
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      22042. .22199
FT      /tag= o
FT      22200. .30359
FT      /tag= p
FT      replace(23340, G)
FT      /tag= q
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(29948, A)
FT      /tag= r
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      30360. .30459
FT      /tag= s
FT      30460. .31475
FT      /tag= t
FT      31476. .31663
FT      /tag= u
FT      31664. .32964
FT      /tag= v
FT      32965. .33087
FT      /tag= w
FT      33088. .34548
FT      /tag= x
FT      replace(33579, C)
FT      /tag= y
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      34549. .34755
FT      /tag= z
FT      34756. .37975
FT      /tag= aa
FT      37976. .38056
FT      /tag= ab
FT      38057. .39552
FT      /tag= ac
FT      39553. .40098
FT      /tag= ad
FT      40099. .46366
FT      /tag= ae
FT      replace(40762, A)
FT      /tag= af
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(40936, C)
FT      /tag= ag
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(45998, G)
FT      /tag= ah
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      46367. .46553
FT      /tag= ai
FT      46554. .49237
FT      /tag= aj
FT      replace(47771, C)
FT      /tag= ak
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(48117, T)
FT      /tag= al
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      49238. .49636
FT      /tag= am
FT      49637. .55445
FT      /tag= an
FT      replace(54563, G)
FT      /tag= ao
```

```
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      55446. .55662
FT      /tag= ap
FT      55663. .62274
FT      /tag= aq
FT      replace(58735, T)
FT      /tag= ar
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(59643, A)
FT      /tag= as
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(61638, G)
FT      /tag= at
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      62275. .62362
FT      /tag= au
FT      replace(63291, G)
FT      /tag= av
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(63463, G)
FT      /tag= aw
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(63636, A)
FT      /tag= ax
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      replace(63998, C)
FT      /tag= ay
FT      /standard_name= "Single nucleotide polymorphism (SNP)"
FT      WO200272764-A2.
FT      19-SEP-2002.
FT      08-MAR-2002; 2002WO-US007156.
FT      13-MAR-2001; 2001US-00804472.
FT      (PEKE ) PE CORP NY.
FT      Shao W, Merkulov GV, Difrancesco V, Beasley EM;
FT      WPI; 2002-759843/82.
FT      DR P-PsDB; AAE29091.
FT      XX
FT      PT New human transporter peptides and polynucleotides encoding the peptides,
FT      useful as models for developing human therapeutic targets, in gene
FT      therapy, or for producing or eliciting an immune response.
FT      PS
FT      Claim 4; Page 93-111, 114pp; English.
FT      XX
FT      CC The invention relates to human transporter polypeptides and nucleic acid
FT      molecules encoding such polypeptides. Sequences of the invention are
FT      useful as models for the development of human therapeutic targets, in the
FT      CC identification of therapeutic proteins and serve as targets for the
FT      development of human therapeutic agents that modulate transporter
FT      activity in cells and tissues that express the transporter. Polypeptides
FT      of the invention can be used to elicit immune response, as reagents in
FT      assays designed to determine the levels of the protein in biological
FT      fluids, as markers for tissue in which the corresponding protein is
FT      preferentially expressed, in the identification of modulators of the
FT      CC peptides and in pharmacogenomic analysis. Nucleic acid molecules of the
FT      CC invention are useful as hybridisation probes, in constructing vectors,
FT      host cells or transgenic animals expressing all or a part of the nucleic
FT      acid, for monitoring the effectiveness of modulating compounds on the
FT      CC expression or activity of the transporter gene in clinical trials or in
FT      treatment regimen, in gene therapy and as antisense constructs to control
FT      CC transporter gene expression in cells, tissue and organisms. The present
FT      sequence is human transporter gene located on chromosome 4
FT      XX
FT      SQ Sequence 65359 BP; 18337 A; 10760 C; 11602 G; 21226 T; 0 U; 3434 Other;
FT      Query Match 6.5%; Score 35.8; DB 6; Length 65359;
FT      Best Local Similarity 49.2%; Pred. No. 23;
```

	Matches	94;	Conservative	0;	Mismatches	97;	Indels	0;	Gaps	0;
Qy	5	CGAGTGTAAATTCCTGATGTAGCTACATTAATTCCTTATTCATCAATCAATTAAGATCAGT	64							
Db	52323	CCAATTTTTTTTTTTTTTTGGTAGAGACAAGGCTTGGCTATGTATCCCAAGCTGGCTTGGAT	52264							
Qy	65	CTTGCGGTACCTCTACGGGCTCTCACCATGCATGCATTCAGATATCCTGTAGACGAT	124							
Db	52263	CTCCTGTGTGCACAACAACCTCCACCTGGCCTCCCATTTAAATAATTTTTTGATGCTA	52204							
Qy	125	GTTATGCAAGAGCCCATTAAGATGAGACAATCTTAATGAACAAGGCTATGACTGTAAA	184							
Db	52203	CTCATTCAGAGGCTTTTAATAAGAACATTACAAATTTTCATTTCTTCATTTTTTAA	52144							
Qy	185	AACAATTTGTA	195							
Db	52143	AAATATTTTTA	52133							

Search completed: September 16, 2006, 14:56:01  
 Job time : 518 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2006, 15:58:20 ; Search time 171 Seconds

(without alignments)  
6072.897 Million cell updates/sec

Title: US-10-815-774-5

Perfect score: 555  
Sequence: 1 ttggcgagtgatctccga.....atgtatccagctgtgattt 555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*  
1: /EMC Celerra SIDS3/ptodata/2/ina/1 COMB.seq:\*  
2: /EMC Celerra SIDS3/ptodata/2/ina/5 COMB.seq:\*  
3: /EMC Celerra SIDS3/ptodata/2/ina/6 COMB.seq:\*  
4: /EMC Celerra SIDS3/ptodata/2/ina/7 COMB.seq:\*  
5: /EMC Celerra SIDS3/ptodata/2/ina/8 COMB.seq:\*  
6: /EMC Celerra SIDS3/ptodata/2/ina/9 COMB.seq:\*  
7: /EMC Celerra SIDS3/ptodata/2/ina/10 COMB.seq:\*  
8: /EMC Celerra SIDS3/ptodata/2/ina/11 COMB.seq:\*  
9: /EMC Celerra SIDS3/ptodata/2/ina/12 COMB.seq:\*  
10: /EMC Celerra SIDS3/ptodata/2/ina/13 COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	555	US-09-727-769A-5	Sequence 5, Appl1
2	555	100.0	1080	US-09-727-769A-7	Sequence 7, Appl1
3	344.2	62.0	555	US-09-324-910-5	Sequence 5, Appl1
4	344.2	62.0	555	US-09-793-495-5	Sequence 5, Appl1
5	344.2	62.0	1080	US-09-324-910-11	Sequence 11, Appl1
6	344.2	62.0	1080	US-09-793-495-11	Sequence 11, Appl1
7	36.4	6.6	474	US-09-621-976-18033	Sequence 18033, A
8	35.8	6.5	101894	US-09-949-016-12005	Sequence 12005, A
9	35.8	6.5	103894	US-09-949-016-14450	Sequence 14450, A
10	35.6	6.4	2489	US-09-141-047-7	Sequence 7, Appl1
11	34.8	6.3	601	US-09-949-016-23364	Sequence 23364, A
12	34.8	6.3	601	US-09-949-016-94061	Sequence 94061, A
13	34.8	6.3	1751	US-09-004-393B-3	Sequence 3, Appl1
14	34.8	6.3	50000	US-09-662-254B-25	Sequence 25, Appl1
15	34.6	6.2	87470	US-09-949-016-15881	Sequence 15881, A
16	34.4	6.2	1072	US-08-956-171E-435	Sequence 435, App
17	34.4	6.2	1072	US-08-781-986A-435	Sequence 435, App
18	34.4	6.2	3510	US-09-487-558B-105	Sequence 105, App
19	34.2	6.2	2019	US-10-009-966C-3	Sequence 3, Appl1
20	34.2	6.2	3666	US-08-682-517-13	Sequence 13, Appl1
21	34.2	6.2	3666	US-08-682-517-14	Sequence 14, Appl1
22	34.2	6.2	4197	US-08-682-517-7	Sequence 7, Appl1
23	34.2	6.2	4197	US-08-682-517-8	Sequence 8, Appl1

C 24	34	6.1	2286	3	US-10-104-047-1871	Sequence 1871, Ap
C 25	34	6.1	56694	3	US-09-949-016-12568	Sequence 12568, A
C 26	34	6.1	56702	3	US-09-949-016-15423	Sequence 15423, A
C 27	33.8	6.1	690	3	US-09-489-039A-6025	Sequence 6025, Ap
C 28	33.6	6.1	1141	3	US-09-806-708B-22	Sequence 22, Appl
C 29	33.6	6.1	4223	3	US-09-541-782-5	Sequence 5, Appl1
C 30	33.6	6.1	4223	3	US-09-723-820-5	Sequence 5, Appl1
C 31	33.6	6.1	4223	3	US-10-270-085-5	Sequence 5, Appl1
C 32	33.6	6.1	36791	3	US-09-949-016-17451	Sequence 17451, A
C 33	33.6	6.1	450395	3	US-09-949-016-15473	Sequence 15473, A
C 34	33.4	6.0	205163	3	US-09-949-016-17009	Sequence 17009, A
C 35	33.2	6.0	601	3	US-09-949-016-96084	Sequence 96084, A
C 36	33.2	6.0	601	3	US-09-949-016-96084	Sequence 96084, A
C 37	33.2	6.0	387902	3	US-09-949-016-14543	Sequence 14543, A
C 38	33.2	6.0	421883	3	US-09-949-016-12557	Sequence 12557, A
C 39	33.2	6.0	1664976	3	US-08-916-421B-1	Sequence 1, Appl1
C 40	33.2	6.0	1664976	3	US-09-692-570-1	Sequence 1, Appl1
C 41	33	5.9	37282	3	US-09-949-016-12440	Sequence 12440, A
C 42	33	5.9	37308	3	US-09-949-016-14162	Sequence 14162, A
C 43	33	5.9	37308	3	US-09-949-016-14163	Sequence 14163, A
C 44	33	5.9	218940	3	US-09-949-016-17539	Sequence 17539, A
C 45	32.8	5.9	601	3	US-09-949-016-39281	Sequence 39281, A

#### ALIGNMENTS

```
RESULT 1
US-09-727-769A-5
; Sequence 5, Application US/09727769A
; Patent No. 6756221
; GENERAL INFORMATION:
; APPLICANT: Amaro Enzyme, Inc.
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, MICROORGANISM PRODUCING THE SA
; FILE REFERENCE: Q62106
; CURRENT APPLICATION NUMBER: US/09/727,769A
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: JP 11-345044
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Cryoseobacterium sp. No. 6756221 9670
US-09-727-769A-5

Query Match      100.0%; Score 555; DB 3; Length 555;
Best Local Similarity 100.0%; Pred. No. 2.1e-159;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGCGAGTGAATTCCTGATGTACCTAATTAAATCTTTATTCATCAATAAAGANT 60
1 TTGGCGAGTGAATTCCTGATGTACCTAATTAAATCTTTATTCATCAATAAAGANT 60
DB 1 TTGGCGAGTGAATTCCTGATGTACCTAATTAAATCTTTATTCATCAATAAAGANT 60
QY 61 CAGTCTTGCGGACTCTTACGCGGCTTCACATGACACACTTGGATATCCCTTAAGC 120
61 CAGTCTTGCGGACTCTTACGCGGCTTCACATGACACACTTGGATATCCCTTAAGC 120
DB 61 CAGTCTTGCGGACTCTTACGCGGCTTCACATGACACACTTGGATATCCCTTAAGC 120
QY 121 GGATGTTATGCAAGAGCCCATTAAGATGAGACAAATCTTAATGAAACAGCGCTATGCT 180
121 GGATGTTATGCAAGAGCCCATTAAGATGAGACAAATCTTAATGAAACAGCGCTATGCT 180
DB 121 GGATGTTATGCAAGAGCCCATTAAGATGAGACAAATCTTAATGAAACAGCGCTATGCT 180
QY 181 GAAAAACAATTTGTATACGAAACCTTAAGGATCAACAGAACTTGTGTGCGCTGG 240
181 GAAAAACAATTTGTATACGAAACCTTAAGGATCAACAGAACTTGTGTGCGCTGG 240
DB 181 GAAAAACAATTTGTATACGAAACCTTAAGGATCAACAGAACTTGTGTGCGCTGG 240
QY 241 AGCTCCACGCTTGCAATTTGGTATGCTTAATAAATGCTCCGAGTAACGGAAGAAAGA 300
241 AGCTCCACGCTTGCAATTTGGTATGCTTAATAAATGCTCCGAGTAACGGAAGAAAGA 300
DB 241 AGCTCCACGCTTGCAATTTGGTATGCTTAATAAATGCTCCGAGTAACGGAAGAAAGA 300
QY 301 ATTATGATCCTTCACACTATTTTCAAGCGGCTCTGTAAAGATGACATGAGAAACGCT 360
301 ATTATGATCCTTCACACTATTTTCAAGCGGCTCTGTAAAGATGACATGAGAAACGCT 360
```

Db 301 ATTATGATCTTCACTAATTTTCAAGCGCTCTGTAACAGATACAGATGAGAAACGCT 360  
Qy 361 TGGCTTAACACCTCTTGGCGATCTGCATCCGTTTCTCTTATGCTAATACTGACGAAAT 420  
Db 361 TGGCTTAACACCTCTTGGCGATCTGCATCCGTTTCTCTTATGCTAATACTGACGAAAT 420  
Qy 421 GTTTATTAACAGAGTCTGATTAATTTCTTACCTGTATGACAACTGATCAATACCAAC 480  
Db 421 GTTTATTAACAGAGTCTGATTAATTTCTTACCTGTATGACAACTGATCAATACCAAC 480  
Qy 481 TGTGTAAGTAAATTTTCACTGCTTCCGATGTTTCTCTTCACTGACCGGATGTA 540  
Db 481 TGTGTAAGTAAATTTTCACTGCTTCCGATGTTTCTCTTCACTGACCGGATGTA 540  
Qy 541 TCCAGCTGTGATTT 555  
Db 541 TCCAGCTGTGATTT 555

RESULT 2  
US-09-727-769A-7  
; Sequence 7, Application US/09727769A  
; Patent No. 6756221  
; GENERAL INFORMATION:  
; APPLICANT: Amamo Enzyme, Inc.  
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, MICROORGANISM PRODUCING THE SAME  
; FILE REFERENCE: 062106  
; CURRENT APPLICATION NUMBER: US/09/727,769A  
; PRIOR FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: JP 11-345044  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 1080  
; TYPE: DNA  
; ORGANISM: Cryseobacterium sp. No. 6756221 9670  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (61)..(1020)  
; OTHER INFORMATION:  
; NAME/KEY: mat peptide  
; LOCATION: (466)..()  
; OTHER INFORMATION:  
; US-09-727-769A-7

Query Match 100.0%; Score 555; DB 3; Length 1080;  
Best Local Similarity 100.0%; Pred. No. 2.7e-159; Mismatches 0; Indels 0; Gaps 0;  
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTTCTTATTCATCAATTAAGAT 60  
Db 466 TTGGCGAGTGAATTCCTGATGTAGCTACATTAATTTCTTATTCATCAATTAAGAT 525  
Qy 61 CAGTCTTGGGATCCTCTACGGGCTCTCACCATGATCACATTCAATATCTCTGTAC 120  
Db 526 CAGTCTTGGGATCCTCTACGGGCTCTCACCATGATCACATTCAATATCTCTGTAC 585  
Qy 121 GGATGTAATGCAAGGCCATAGATGACAAATCTTAATGAACAACGGCTATGATGT 180  
Db 586 GGATGTAATGCAAGGCCATAGATGACAAATCTTAATGAACAACGGCTATGATGT 645  
Qy 181 GAAAAACAATTTGTATACGAAACCTAAAGGATCAACAGGAACCTTGTGTGGCGTGG 240  
Db 646 GAAAAACAATTTGTATACGAAACCTAAAGGATCAACAGGAACCTTGTGTGGCGTGG 705  
Qy 241 AGTACCAAGCTTGAATATTTGTATAGCTATATAAATGCTTCCGAGTACGGAAGAA 300  
Db 706 AGTACCAAGCTTGAATATTTGTATAGCTATATAAATGCTTCCGAGTACGGAAGAA 765  
Qy 301 ATTATGATCTTCACTAATTTTCAAGCGCTCTGTAACAGATACAGATGAGAAACGCT 360

Db 766 ATTATGATCTTCACTAATTTTCAAGCGCTCTGTAACAGATACAGATGAGAAACGCT 825  
Qy 361 TGGCTTAACACCTCTTGGCGATCTGCATCCGTTTCTCTTATGCTAATACTGACGAAAT 420  
Db 826 TGGCTTAACACCTCTTGGCGATCTGCATCCGTTTCTCTTATGCTAATACTGACGAAAT 885  
Qy 421 GTTTATTAACAGAGTCTGATTAATTTCTTACCTGTATGACAACTGATCAATACCAAC 480  
Db 886 GTTTATTAACAGAGTCTGATTAATTTCTTACCTGTATGACAACTGATCAATACCAAC 945  
Qy 481 TGTGTAAGTAAATTTTCACTGCTTCCGATGTTTCTCTTCACTGACCGGATGTA 540  
Db 946 TGTGTAAGTAAATTTTCACTGCTTCCGATGTTTCTCTTCACTGACCGGATGTA 1005  
Qy 541 TCCAGCTGTGATTT 555  
Db 1006 TCCAGCTGTGATTT 1020

RESULT 3  
US-09-324-910-5  
; Sequence 5, Application US/09324910  
; Patent No. 6251651  
; GENERAL INFORMATION:  
; APPLICANT: Matsura, Akira  
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE  
; FILE REFERENCE: Q-54362  
; CURRENT APPLICATION NUMBER: US/09/324,910  
; PRIOR FILING DATE: 1999-06-03  
; EARLIER APPLICATION NUMBER: HEI-10-173940  
; EARLIER FILING DATE: 1998-06-04  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 555  
; TYPE: DNA  
; ORGANISM: Cryseobacterium gleum  
; US-09-324-910-5

Query Match 62.0%; Score 344.2; DB 3; Length 555;  
Best Local Similarity 76.7%; Pred. No. 5.1e-95; Mismatches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;  
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;  
Qy 7 AGTGAATTCCTGATGTAGCTACATTAATTTCTTATTCATCAATTAAGATCAGTCT 66  
Db 7 AGTGAATTCCTGATGTAGCTACATTAATTTCTTATTCATCAATTAAGATCAGTCT 66  
Qy 67 TGGGATCCTTAACGGGCTCTCACCATGATCACATTAAGATCAGTCTGATGT 126  
Db 67 TGGGATCCTTAACGGGCTCTCACCATGATCACATTAAGATCAGTCTGATGT 126  
Qy 127 TATGCAAGGCCATTAAGATGAGACAAATCTTAATGAACAACGGCTATGATGTA 186  
Db 127 TATGCAAGGCCATTAAGATGAGACAAATCTTAATGAACAACGGCTATGATGTA 186  
Qy 187 CAATTTGTATACGAAACCTAAAGGATCAACAGAACTTGTGTGTGGCGTAC 246  
Db 187 CAATTTGTATACGAAACCTAAAGGATCAACAGAACTTGTGTGTGGCGTAC 246  
Qy 247 CAGTGTCAATATTTGTATGCTTAATAATGCTTCCGAGTACGGAAGAAATTAAT 306  
Db 247 CAGTGTCAATATTTGTATGCTTAATAATGCTTCCGAGTACGGAAGAAATTAAT 306  
Qy 307 GATCTTCACTATTTTCAAGCGGCTCTGTAACAGATACAGATGAGAAACGCTTGC 366  
Db 307 GATCTTCACTATTTTCAAGCGGCTCTGTAACAGATACAGATGAGAAACGCTTGC 366  
Qy 367 AACACCTCTGCGGATCTGCATCCGTTTCTCTTATGCTAATACTGACGAAATGTTAT 426  
Db 367 AACACCTCTGCGGATCTGCATCCGTTTCTCTTATGCTAATACTGACGAAATGTTAT 426





; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12005  
 ; LENGTH: 101894  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(101894)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-12005

Query Match 6.5%; Score 35.8; DB 3; Length 101894;  
 Best Local Similarity 49.2%; Pred. No. 6.9;  
 Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 5 CGAGGTAAATTCCTGATGAGTACATTAATTTCTTATTCATCAATTAAGATCACT 64  
 DB 41497 CCAATTTTTTTTTTTTGGTAGACAAAGGCTTGCTATGTTACCAAGCTGCTTGTAT 41438  
 QY 65 CTGGGGTACCTCTACGGGCTCTCCACATGCATCATTGAGATCTCTGAGAGGAT 124  
 DB 41437 CTCCTGGTCCAAACAAAGCTCCACCTCGGCTCCCATTTAAATATTTTGTATGCTA 41378  
 QY 125 GTTATGCAAGAGCCCATTAAGATGAGCAATCTTAATGAACAAGGCTATGACTGTGAAA 184  
 DB 41377 CTCATTCAGAGGCTTTTAAATAGGACATTACATTTTCTTCATTTTAA 41318  
 QY 185 AACATTTGTA 195  
 DB 41317 AATATTTTTTA 41307

RESULT 9  
 US-09-949-016-14450/c  
 ; Sequence 14450, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 14450  
 ; LENGTH: 103894  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(103894)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-14450

Query Match 6.5%; Score 35.8; DB 3; Length 103894;  
 Best Local Similarity 49.2%; Pred. No. 7;  
 Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
 QY 5 CGAGGTAAATTCCTGATGAGTACATTAATTTCTTATTCATCAATTAAGATCACT 64

DB 91497 CCAATTTTTTTTTTTTGGTAGACAAAGGCTTCTATGTTACCAAGCTGCTTGTAT 91438  
 QY 65 CTGGGGTACCTCTACGGGCTCTCCACATGCATCATTGAGATCTCTGAGAGGAT 124  
 DB 91437 CTCCTGGTCCAAACAAAGCTCCACCTCGGCTCCCATTTAAATATTTTGTATGCTA 91378  
 QY 125 GTTATGCAAGAGCCCATTAAGATGAGCAATCTTAATGAACAAGGCTATGACTGTGAAA 184  
 DB 91377 CTCATTCAGAGGCTTTTAAATAGGACATTACATTTTCTTCATTTTAA 91318  
 QY 185 AACATTTGTA 195  
 DB 91317 AATATTTTTTA 91307

RESULT 10  
 US-09-141-047-7/c  
 ; Sequence 7, Application US/09141047A  
 ; Patent No. 6043085  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, David H.  
 ; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic  
 ; FILE REFERENCE: D6143  
 ; CURRENT APPLICATION NUMBER: US/09/141,047A  
 ; CURRENT FILING DATE: 1998-08-27  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SEQ ID NO 7  
 ; LENGTH: 2489  
 ; TYPE: DNA  
 ; ORGANISM: Ehrlichia canis  
 ; FEATURE:  
 ; LOCATION: -340..2149  
 ; OTHER INFORMATION: Nucleotide sequence of gene encoding 120kDa  
 ; OTHER INFORMATION: Immunoreactive protein  
 US-09-141-047-7

Query Match 6.4%; Score 35.6; DB 3; Length 2489;  
 Best Local Similarity 45.9%; Pred. No. 1.7;  
 Matches 122; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 276 TGCTTCGGAGTACGAGAAAAAGATTTATGATCTCTACTATTTTCAAGGGTCTGT 335  
 DB 1843 TTCTTCCTTACTAGTTTCAAGATCTTTTCTCAACTTCACTGATGATGTTTCACT 1784  
 QY 336 AACAGATACAGATGAGAAAGCTTGGTTTAACTCTTGGGATCTGATCCGTTTC 395  
 DB 1783 ACCATCTACAGAGGTTCAATCTTCGCTTTTAACTTCAAGAGTATTTTCTTACT 1724  
 QY 396 CTCCTATGCTAATATGAGAAATGTTTATTAAGAAAGTCTTATTAATTTCTTACTGTA 455  
 DB 1723 AGTTTCAATCTTTTCTTCAACTTCACTGATGATGATGTTTCACTACATCTACACG 1664  
 QY 456 TGACAAATCTGATCAATACCAACTGTGTACTGTAATTTTCACTGCTTTCCGAGT 515  
 DB 1663 AGTTGCAATCTTCTGCTTAACTTCAAGATATTTTCTTACTAGTGTGATAC 1604  
 QY 516 TTCTCTTACCTGACCGGATGAT 541  
 DB 1603 TTTTCTCACTTCACTGATGAT 1578

RESULT 11  
 US-09-949-016-23364/c  
 ; Sequence 23364, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016

```

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23364
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-23364

Query Match
Best Local Similarity 50.0%; Pred. No. 1.7;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 22 GTAGCTACATTAATTTCTTATTCATCAATTAAGATGAGTCTGGCTAGCTTACG 81
DB 592 GTAGAGACAAAGGTCTTGATGTACCCAAAGCTGTGATCTCTGTGACCAACAA 533
QY 82 GCGTCTCACCATCATCATTCAGATATCTGTAGACGAGATGTATGCAAGGCCAT 141
DB 532 ACCTCCACCTGCGCTCCCATTTAAANATTTTGATGCTACTCATTCAGGCGCTT 473
QY 142 AAGATGAGCAATCTTAATGAACAACGGCTATGACTGTGAAAAACAATTTGTA 195
DB 472 TAAATAGGAACATTAATTTCTTCAATTTCTTCAATTTTAAAAATATTTTAA 419

RESULT 12
US-09-949-016-94061/c
; Sequence 94061, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTUR, J. Craig et al.
; TITLE OF INVENTION: WITH HUMAN DISIN IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94061
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-94061

Query Match
Best Local Similarity 6.3%; Score 34.8; DB 3; Length 601;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 22 GTAGCTACATTAATTTCTTATTCATCAATTAAGATGAGTCTGGCTAGCTTACG 81
DB 592 GTAGAGACAAAGGTCTTGATGTACCCAAAGCTGTGATCTCTGTGACCAACAA 533
QY 82 GCGTCTCACCATCATCATTCAGATATCTGTAGACGAGATGTATGCAAGGCCAT 141
DB 532 ACCTCCACCTGCGCTCCCATTTAAANATTTTGATGCTACTCATTCAGGCGCTT 473
QY 142 AAGATGAGCAATCTTAATGAACAACGGCTATGACTGTGAAAAACAATTTGTA 195
DB 472 TAAATAGGAACATTAATTTCTTCAATTTCTTCAATTTTAAAAATATTTTAA 419
```

```

RESULT 13
US-09-004-393B-3
; Sequence 3, Application US/09004393B
; Patent No. 6310271
; GENERAL INFORMATION:
; APPLICANT: Hanson D., Andrew
; APPLICANT: Rathnasabapathi, Bala
; APPLICANT: Burnett, Michael
; TITLE OF INVENTION: Polynucleotides Encoding Choline Monooxygenase and
; TITLE OF INVENTION: Plants Transformed Therewith
; FILE REFERENCE: UP-162
; CURRENT APPLICATION NUMBER: US/09/004,393B
; CURRENT FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: 60/035,147
; PRIOR FILING DATE: 1997-01-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1751
; TYPE: DNA
; ORGANISM: Beta vulgaris
US-09-004-393B-3

Query Match
Best Local Similarity 6.3%; Score 34.8; DB 3; Length 1751;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 274 AATGCTTCGGAGTAAACGAAAAAATTAATGATCTTCACTATTTTCAAGCGTCT 333
DB 643 AAGCCAGCAAGCACTAAACAAATTTGATCTTAAAGAACTTGCGCTTGACACC 702
QY 334 GTAAAGATATACAGATGAGAAACGCTTGTAAACACCTCTTGCGATCTGCATCCGT 393
DB 703 CTAAAGTTGCAAGATGGGCCCCATTCATTTATTCAGCTTGACCGATCTTAATGCT 762
QY 394 TCCCTTATGCTAATACCTCAGAAATGTTTATTAACAGAGTCTAGTAATTTTACCTG 453
DB 763 AATCGTATGTTGAAACAGAGTGTATATCTGCAAGATGTAAAGGCCCATGCT 822
QY 454 TATGACAAATCTGATCAATACCA 479
DB 823 TTTGATCTTAATCTAAAGTTCAACCA 848

RESULT 14
US-09-662-254B-25
; Sequence 25, Application US/09662254B
; Patent No. 6931145
; GENERAL INFORMATION:
; APPLICANT: Meyer, Richard W.
; APPLICANT: Li, Yi
; APPLICANT: Bowen, Alison Louise
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
; FILE REFERENCE: UP-221C1X1
; CURRENT APPLICATION NUMBER: US/09/662,254B
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/086,651
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Version 3.1
; SEQ ID NO 25
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amsacta moorei entomopoxvirus
US-09-662-254B-25

Query Match
Best Local Similarity 6.3%; Score 34.8; DB 3; Length 50000;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```



QY	116	TAGACGGATGTTATGCAAGAGCCCATTAAGTGGACAAATCTTAAGAACAACGGGTAG	175
Db	28661	TTGGTGGATATTTTGGTTTAAAGAACAAATTAATACATTAGCTATGAATTAATTCGAAT	28720
QY	176	ACTGTGAAAAACAATTGTATTACGSAACCTTAAGGCATCAACAGAACCTTGCTGTGTG	235
Db	28721	TAAATGAAAACTATCTATTTTCCATACATTAACATGTCACCTGTAAATTAATTTTGAAG	28780
QY	236	CGTGAAGCTACCAAGCTGCATATTGGTAAAGCTATAA	273
Db	28781	GATGTAGTAACCCAGATGCTATATTGATAGATAAAAA	28818

```

RESULT 15
US-09-949-016-15881/C
Sequence 15881, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT FILING DATE: 2000-04-14
PRIORITY FILING DATE: 2000-10-20
PRIORITY FILING DATE: 2000-10-03
PRIORITY FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15881
LENGTH: 87470
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)---(87470)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15881

```

[illegible]

Search completed: September 16, 2006, 17:06:13  
Job time : 177 secs

This Page Blank (uspto)



```
Db      121  |||||
Qy      181  |||||
Db      181  |||||
Qy      241  |||||
Db      241  |||||
Qy      301  |||||
Db      301  |||||
Qy      361  |||||
Db      361  |||||
Qy      421  |||||
Db      421  |||||
Qy      481  |||||
Db      481  |||||
Qy      541  |||||
Db      541  |||||
```

```
RESULT 2
US-10-815-774-7
; Sequence 7, Application US/10815774
; Publication No. US20040175799A1
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, MICROORGANISM PRODUCING THE SAME
; FILE REFERENCE: Q80844
; CURRENT APPLICATION NUMBER: US/10/815,774
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US/09/727,769
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US/09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: JP Hei. 11-345044
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP Hei. 10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Cryoseobacterium sp. No. 9670
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)..(1020)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (466)..()
US-10-815-774-7
```

```
Query Match      100.0%; Score 555; DB 9; Length 1080;
Best Local Similarity 100.0%; Pred. No. 1.6e-153;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1  |||||
Db      466  |||||
```

```
Qy      61  |||||
Db      526  |||||
Qy      121  |||||
Db      586  |||||
Qy      181  |||||
Db      646  |||||
Qy      241  |||||
Db      706  |||||
Qy      301  |||||
Db      766  |||||
Qy      361  |||||
Db      826  |||||
Qy      421  |||||
Db      886  |||||
Qy      481  |||||
Db      946  |||||
Qy      541  |||||
Db      1006 |||||
```

```
RESULT 3
US-09-793-495-5
; Sequence 5, Application US/09793495
; Publication No. US20040072318A1
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE SAME
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/793,495
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: HEI-10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Cryoseobacterium gleum
US-09-793-495-5
```

```
Query Match      62.0%; Score 344.2; DB 3; Length 555;
Best Local Similarity 76.7%; Pred. No. 4.1e-91;
Matches 421; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
```

```
Qy      7  |||||
Db      7  |||||
Qy      67  |||||
Db      67  |||||
```





Db 151 GCTACCTGTTATATAGGGGTTAGGCTAGAGTACCCAAAGGAAAAAGGTAATTG 210  
Qy 302 TTAATGATCCTTACATATTTTCAAGCGTCTGTACAG 340  
Db 211 CTAAAGAAATTCACAAAAATTAAGTTGCTTTAAAAAG 249

## RESULT 8

US-09-925-065A-874704  
; Sequence 874704, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 874704  
; LENGTH: 622  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-874704

Query Match Best Local Similarity 6.8%; Score 37.6; DB 4; Length 622;  
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 146 TGAGCAAACTTAAATGAACAACGGCTATGACTGTGAAAAACAATTTGTATACGGAACC 205  
Db 462 TGAATATATGTTAAATTAATTAATGCTATTAATATGAACACCATTAATATAAAAAAT 521  
Qy 206 TAAAGCATCAACAGAACTTCTGTGTGCGGTGAGCTA 245  
Db 522 TAAAAAGACCAACAGAAATTTGGAGATGCTTTAAGCTA 561

## RESULT 9

US-09-925-065A-874705  
; Sequence 874705, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 874705  
; LENGTH: 622  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-874705

Query Match Best Local Similarity 6.8%; Score 37.6; DB 4; Length 622;  
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 146 TGAGCAAACTTAAATGAACAACGGCTATGACTGTGAAAAACAATTTGTATACGGAACC 205  
Db 462 TGAATATATGTTAAATTAATTAATGCTATTAATATGAACACCATTAATATAAAAAAT 521  
Qy 206 TAAAGCATCAACAGAACTTCTGTGTGCGGTGAGCTA 245  
Db 522 TAAAAAGACCAACAGAAATTTGGAGATGCTTTAAGCTA 561

## RESULT 10

US-09-925-065A-874706  
; Sequence 874706, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 874706  
; LENGTH: 622  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-874706

Query Match Best Local Similarity 6.8%; Score 37.6; DB 4; Length 622;  
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 146 TGAGCAAACTTAAATGAACAACGGCTATGACTGTGAAAAACAATTTGTATACGGAACC 205  
Db 462 TGAATATATGTTAAATTAATTAATGCTATTAATATGAACACCATTAATATAAAAAAT 521  
Qy 206 TAAAGCATCAACAGAACTTCTGTGTGCGGTGAGCTA 245  
Db 522 TAAAAAGACCAACAGAAATTTGGAGATGCTTTAAGCTA 561

## RESULT 11

US-09-925-065A-874707  
; Sequence 874707, Application US/09925065A  
; Publication No. US20040181048A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096

```

; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 874707
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-874707

Query Match
Best Local Similarity 61.0%; Score 37.6; DB 4; Length 622;
Best Local Similarity 61.0%; Pred. No. 4.3;
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

146 TGAGACAAATCTTAATGAAACAGCGCTATGCTGTGAAAAACAATTGTATACGGAACC 205
462 TGAATATATGTTAATTAATTAATGCGTATTAATATGAAACACCATTAATATAAAAAAAT 521
DB 206 TAAAGCATCAACAGAACTTCTGTGTGCGCTGAGCTA 245
522 TAAAAAGCAACCAAGATATTGTGAGATGCTTTGTACTA 561

RESULT 12
US-09-925-065A-874704
; Sequence 874704, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 874704
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-874704

Query Match
Best Local Similarity 61.0%; Score 37.6; DB 5; Length 622;
Best Local Similarity 61.0%; Pred. No. 4.3;
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

146 TGAGACAAATCTTAATGAAACAGCGCTATGCTGTGAAAAACAATTGTATACGGAACC 205
462 TGAATATATGTTAATTAATTAATGCGTATTAATATGAAACACCATTAATATAAAAAAAT 521
DB 206 TAAAGCATCAACAGAACTTCTGTGTGCGCTGAGCTA 245
522 TAAAAAGCAACCAAGATATTGTGAGATGCTTTGTACTA 561

RESULT 13
```

```

US-09-925-065A-874705
; Sequence 874705, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 874705
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-874705

Query Match
Best Local Similarity 61.0%; Score 37.6; DB 5; Length 622;
Best Local Similarity 61.0%; Pred. No. 4.3;
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

146 TGAGACAAATCTTAATGAAACAGCGCTATGCTGTGAAAAACAATTGTATACGGAACC 205
462 TGAATATATGTTAATTAATTAATGCGTATTAATATGAAACACCATTAATATAAAAAAAT 521
DB 206 TAAAGCATCAACAGAACTTCTGTGTGCGCTGAGCTA 245
522 TAAAAAGCAACCAAGATATTGTGAGATGCTTTGTACTA 561

RESULT 14
US-09-925-065A-874706
; Sequence 874706, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 874706
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-874706

Query Match
Best Local Similarity 61.0%; Score 37.6; DB 5; Length 622;
Best Local Similarity 61.0%; Pred. No. 4.3;
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```



Qy	Db	Qy	Db
146	462	206	522
TGAACAAATCTTATATGAAACAACGGCTATGCTGTGAAAAACATTTGTATACGGAAAC	TGAATATATGTTAATATAATTTATGGCTATTTAATATGAAACCCCATTTAATATTAATAAATAAT	TAAAGCATCAACAGAACTGCTGTGTGGCTGGAGACTA	TAAAAAGACAAACGAATATTTTGGAAATGCTTGTATACCTA
205	521	245	561

RESULT 15

```

US-09-925-065A-874707
: Sequence 874707, Application US/09925065A
: Publication No. US20050228172A9
: GENERAL INFORMATION:
:   APPLICANT: Wang, David G.
:   TITLE OF INVENTION: Identification and Mapping of Single
:   TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
:   FILE REFERENCE: 108827.135
:   CURRENT APPLICATION NUMBER: US/09/925,065A
:   CURRENT FILING DATE: 2001-08-08
:   PRIOR APPLICATION NUMBER: US 60/243,096
:   PRIOR FILING DATE: 2000-10-24
:   PRIOR APPLICATION NUMBER: US 60/252,147
:   PRIOR FILING DATE: 2000-11-20
:   PRIOR APPLICATION NUMBER: US 60/250,092
:   PRIOR FILING DATE: 2000-11-30
:   PRIOR APPLICATION NUMBER: US 60/261,766
:   PRIOR FILING DATE: 2001-01-16
:   PRIOR APPLICATION NUMBER: US 60/289,846
:   PRIOR FILING DATE: 2001-05-09
:   NUMBER OF SEQ ID NOS: 957086
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 874707
:   LENGTH: 622
:   TYPE: DNA
:   ORGANISM: Homo sapiens
US-09-925-065A-874707

```

Query Match	Score	DB	Length
6.8%	37.6	DB 5	622

Best Local Similarity 61.0%; Pred. No. 4.3;  
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0

Dy      146 TGAGCAAATCTTAATGAACACGCGTATGCATGTGAAAAACAATTGGTATACGGAAACC 205

Db      462 TGAATATAGTTAAATAAATTATGCGTATTATATGAACACACTTATATATAAAAAAAAAAT 521

QY		206 TAAAGCATCAACAGAACTTGTGTCGGCGTGGAAGCTA	245
Db		522 TAAAAAGAGCAACAGATATTGGAGATGCTTTGTAAGCTA	561

Search completed: September 16, 2006, 21:50:50  
Job time : 1126 secs

This Page Blank (uspto)

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using bw model

Run on: September 16, 2006, 18:31:40 ; Search time 193 Seconds  
(without alignments)  
5043.212 Million cell updates/sec

Title: US-10-815-774-5  
Sequence: 1 ttcgcgcagtgtaactcctga.....atgatccagctggtatc 555

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2330496 seqs, 876882855 residues.

Total number of hits satisfying chosen parameters: 4660992

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA\_New.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US09\_NEW\_PUB\_seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US06\_NEW\_PUB\_seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US07\_NEW\_PUB\_seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US08\_NEW\_PUB\_seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/PCT\_NEW\_PUB\_seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US10\_NEW\_PUB\_seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB\_seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB\_seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US11\_NEW\_PUB\_seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/1/pubpna/US60\_NEW\_PUB\_seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.8	7.2	634888	6 US-10-533-365-1	Sequence 1, Appl
2	34.4	6.2	777	6 US-10-471-571A-2813	Sequence 2813, Ap
3	34.4	6.2	3510	8 US-11-217-529-191047	Sequence 191047,
4	34	6.1	822	8 US-11-266-748A-40049	Sequence 40049, A
5	34	6.1	822	8 US-11-266-748A-214137	Sequence 214137,
6	34	6.1	1744	9 US-11-218-305-11243	Sequence 11243, A
7	34	6.1	4620	8 US-11-266-748A-27304	Sequence 27304, A
8	34	6.1	4620	8 US-11-266-748A-31292	Sequence 31292, A
9	33.6	6.1	349	7 US-11-314-834-716	Sequence 716, App
10	33.6	6.1	339234	6 US-10-669-920-1264	Sequence 1264, Ap
11	33.4	6.0	1000	8 US-11-266-748A-207141	Sequence 207141,
12	33.4	6.0	4647455	6 US-10-641-321-205	Sequence 205, App
13	33.2	6.0	397	8 US-11-266-748A-411280	Sequence 411280,
14	33	5.9	1000	8 US-11-266-748A-206397	Sequence 206397,
15	33	5.9	1467	6 US-10-953-349-36486	Sequence 36486, A
16	33	5.9	1602	9 US-11-218-305-16633	Sequence 16632, A
17	33	5.9	2582	9 US-11-218-305-16633	Sequence 16633, A
18	33	5.9	100137	6 US-10-540-898-620	Sequence 620, App
19	33	5.9	150000	8 US-11-266-748A-23591	Sequence 23591, A
20	33	5.9	242825	8 US-11-266-748A-33745	Sequence 32745, A
21	32.8	5.9	612	8 US-11-266-748A-255476	Sequence 255476,
22	32.8	5.9	612	8 US-11-266-748A-315993	Sequence 315993,
23	32.8	5.9	4581	8 US-11-217-529-79437	Sequence 79437, A

24	32.8	5.9	5504	7 US-11-329-884-1	Sequence 1, Appl
25	32.8	5.9	6124	7 US-11-329-884-21	Sequence 21, Appl
26	32.8	5.9	6133	7 US-11-329-884-2	Sequence 2, Appl
27	32.8	5.9	6133	7 US-11-363-149-38	Sequence 38, Appl
28	32.8	5.9	6133	7 US-11-363-151-38	Sequence 38, Appl
29	32.8	5.9	6133	8 US-11-266-748A-29518	Sequence 29518, A
30	32.8	5.9	6133	8 US-11-266-748A-56391	Sequence 56391, A
31	32.8	5.9	6228	7 US-11-329-884-23	Sequence 23, Appl
32	32.6	5.9	201	6 US-10-284-444-30584	Sequence 30584, A
33	32.6	5.9	21909	8 US-11-266-748A-24012	Sequence 24012, A
34	32.4	5.8	555	7 US-11-292-078-2805	Sequence 2805, Ap
35	32.4	5.8	1000	8 US-11-266-748A-202893	Sequence 202893,
36	32.4	5.8	3715	8 US-11-266-748A-78432	Sequence 78432, A
37	32.4	5.8	3715	8 US-11-266-748A-110231	Sequence 110231,
38	32.4	5.8	3715	8 US-11-266-748A-131243	Sequence 131243,
39	32.4	5.8	70665	6 US-10-505-928-596	Sequence 596, App
40	32.4	5.8	523543	6 US-10-540-898-308	Sequence 308, App
41	32.2	5.8	4337	8 US-11-293-697-1232	Sequence 1232, Ap
42	32	5.8	1000	8 US-11-266-748A-199526	Sequence 199526,
43	31.8	5.7	1029	8 US-11-217-529-78011	Sequence 78011, A
44	31.8	5.7	2226	6 US-10-471-571A-1893	Sequence 1893, Ap
45	31.8	5.7	2676	8 US-11-216-545-1960	Sequence 1960, Ap

## ALIGNMENTS

RESULT 1  
US-10-533-365-1  
; Sequence 1, Application US/10533365  
; Publication No. US2006014162A1  
; GENERAL INFORMATION:  
; APPLICANT: Reynoldsclif, Inga  
; APPLICANT: Gulcher, Jeffrey R.  
; APPLICANT: Grant, Struan F.  
; APPLICANT: Thorleifsson, Gudmar  
; TITLE OF INVENTION: Human Type II Diabetes Gene - Silt-3  
; FILE REFERENCE: 2345.2046-007  
; CURRENT APPLICATION NUMBER: US/10/533,365  
; CURRENT FILING DATE: 2005-04-29  
; PRIOR APPLICATION NUMBER: PCT/US03/34801  
; PRIOR FILING DATE: 2003-10-31  
; PRIOR APPLICATION NUMBER: 60/423,541  
; PRIOR FILING DATE: 2002-11-01  
; NUMBER OF SEQ ID NOS: 236  
; SOFTWARE: PaacSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 634888  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-10-533-365-1

Query Match  
Best Local Similarity 49.3%; Pred. No. 0.61;  
Matches 104; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY	18	TCATGAGCTAATTAATTTCTTATTCATGAATTAAGATTCAGTTCGGTACCC	77
DB	385208	TGCTTTGTTTAACATCAATCTATACACAGTTCATATGACATTTACTGTGAGCC	385267
QY	78	TACGCGCTCCATCAGATGATCATTGATATCTGTGAGCGATGTATGACAGAC	137
DB	385268	TTTTTGTTCACAAACATGCTGCTTCATACAGAGCTCTGACATATGATGATGAC	385327
QY	138	CCATTAAGATGACAAATCTTAATGAACAACGCTATGACTGTGAAAAAATTTGTATA	197
DB	385328	ACATTAATATGCTAGTTCATTAATTAAGATGCTGTATTAATTAATCAAGTGAT	385387
QY	198	CGGAACCTTAAGCATCAACGAACCTGC	228
DB	385388	TCCAAAACGATTCACAAAAAAGAAATGC	385418

```
RESULT 2
US-10-471-571A-2813/c
; Sequence 2813, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWIn99, version 1.03
; SEQ ID NO 2813
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-471-571A-2813

Query Match
Best Local Similarity 6.2%; Score 34.4; DB 6; Length 777;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 329 GTCTGTAAACAGATACGCGTGGAGAAACGCTTACACTCTTGGCGATCGCAT 388
DB 247 GTTCTGAATAATATATACGAGTGGCTTACGCTTACCACTCTTGTGTTAGCAT 188

QY 389 CCGTTTCTCTTAATGCTAATATCTGCAGAAAT 420
DB 187 CAATGTGCAATTTACGTAATTTTTCAGTTACT 156

RESULT 3
US-11-217-529-191047
; Sequence 191047, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; PRIOR FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 191047
; LENGTH: 3510
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-191047

Query Match
Best Local Similarity 6.2%; Score 34.4; DB 8; Length 3510;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 292 GAAAAAAGATTTATGATCTTCACTATTTTCAAGCGTCTGTACAGATACAGCATGG 351
DB 3148 GAAATAGTGCACGAAGAGCTTCATTCGTTTCAGAGGCTGTGAGACAAATTCATAGC 3207

QY 352 AGAAACGCTTGCTTAACACCTCTGGAGATCGATCGTTTCTTATGCTAATCT 411
DB 3208 AGAAACAACAATTTTCTCAACTCTGGACTTCTTATTTGAACACATACACCACT 3267

QY 412 GCAGGAATGTTTATTACAGAGTCTAGTAATCT 447
```

```
DB 3268 ACAAGCAGCGCTGTAATAATAGCAGCAATGTTTCT 3303

RESULT 4
US-11-266-748A-40049/c
; Sequence 40049, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcriptome Microarray Technology and
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
; PRIOR FILING DATE: 2005-11-03
; PRIOR APPLICATION NUMBER: EP 04105479.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105482.6
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105483.4
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105507.0
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105485.9
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: EP 04105484.2
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/662,276
; PRIOR FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/700,293
; PRIOR FILING DATE: 2005-07-18
; NUMBER OF SEQ ID NOS: 483996
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 40049
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (806)..(806)
; OTHER INFORMATION: n is a, c, g, or t
US-11-266-748A-40049

Query Match
Best Local Similarity 6.1%; Score 34; DB 8; Length 822;
Matches 70; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 102 ATTGAGATATCTCTGTAGACGATGTTATGCAAGAGCCATTAAGTGAACAATCTTAT 161
DB 552 ATTCATTTTTCACGGAGAAATATACATAGCAAGGTAAACAGTGAAGCCAGACCTAAT 493

QY 162 GAACAAGCGTATGACTGTGAATAACAATTTGTATACGGAACCTTAAGGATCAACAG 221
DB 492 CAATTAATGCTTACACCTGAAATATATGTAAAGTCTTAATGCCAATAGTACTAACGA 433

QY 222 AACTTGCTGT 231
DB 432 GACTTGATAT 423

RESULT 5
US-11-266-748A-214137/c
; Sequence 214137, Application US/11266748A
; Publication No. US20060134663A1
; GENERAL INFORMATION:
; APPLICANT: Harkin, Paul
; APPLICANT: Johnston, Patrick
; APPLICANT: Mulligan, Karl
; TITLE OF INVENTION: Transcriptome Microarray Technology and
; FILE REFERENCE: 55815-0102 (319189)
; CURRENT APPLICATION NUMBER: US/11/266,748A
```



```
APPLICANT: Harkin, Paul
APPLICANT: Johnston, Patrick
APPLICANT: Mulligan, Karl
TITLE OF INVENTION: Transcriptome Microarray Technology and
FILE REFERENCE: 55815-0102 (319189)
CURRENT APPLICATION NUMBER: US/11/266,748A
PRIOR FILING DATE: 2005-11-03
PRIOR APPLICATION NUMBER: EP 04105479.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105482.6
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105483.4
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105507.0
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105485.9
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105484.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: US 60/662,276
PRIOR FILING DATE: 2005-03-14
PRIOR APPLICATION NUMBER: US 60/700,293
PRIOR FILING DATE: 2005-07-18
NUMBER OF SEQ ID NOS: 48396
SOFTWARE: PatentIn version 3.3
SEQ ID NO 31292
LENGTH: 4620
TYPE: DNA
ORGANISM: Homo Sapiens
US-11-266-748A-31292

Query Match
Best Local Similarity 5.1%; Score 34; DB 8; Length 4620;
Matches 70; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 102 ATTGAGATATCTGTAGACGATGTTATGCAAGGCCCATTAAGTGAACAATCTTAAT 161
    |||||
Db 2176 ATTCATTTTACCGGAGAAAATACATAGCAAGGATTAACAGTGAGCCAAAGCTTAAT 2117

QY 162 GAACACGGCTATGACTGTGAAAAACAATTTGTATACGAAACCTTAAGGATCAAGG 221
    |||||
Db 2116 CAATATATGCTTAGCCTGAAATATGTTAAGTCTTCTTAATGCAATAGTACTAACA 2057

QY 222 AACTGCTGT 231
    |||||
Db 2056 GACTTGATAT 2047

RESULT 9
US-11-314-834-716
Sequence 716, Application US/11314834
Publication No. US20060195942A1
GENERAL INFORMATION:
APPLICANT: Fincher, Karen L.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 16517.360 - 38-21(51770)C/US
CURRENT APPLICATION NUMBER: US/11/314,834
CURRENT FILING DATE: 2005-12-22
PRIOR APPLICATION NUMBER: US 09/732,627
PRIOR FILING DATE: 2000-12-08
NUMBER OF SEQ ID NOS: 4930
SEQ ID NO 716
LENGTH: 349
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: Clone ID: LIB3493-050-PI-M1-C3
US-11-314-834-716

Query Match
Best Local Similarity 6.1%; Score 33.6; DB 7; Length 349;
Matches 55.0%; Pred. No. 1.2;
```

```
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 326 GCGTCTGTGACATACGATGCAAGAAACCTTCCGTACACCTCTGGGATCTG 385
    |||||
Db 22 GGGCTGGCTAGAGAAATGCTTGAGCTTTGGCCCATGAACCATCTCAGGCTGCTG 81

QY 386 CATCGTTTCTCTTATGCTAATATGCAAGAAATGTTATTAAGAAAGCTTAGTAAT 445
    |||||
Db 82 CATCTGCTCATCTGTTCTTCCAAATGTTGAGAAAGGTTCCCAAAATAGTGTATT 141

RESULT 10
US-10-669-920-1264
Sequence 1264, Application US/10669920
Publication No. US20060194265A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
FILE REFERENCE: 20366-066001
CURRENT APPLICATION NUMBER: US/10/669,920
PRIOR FILING DATE: 2003-09-23
PRIOR APPLICATION NUMBER: US 10/004,113
PRIOR FILING DATE: 2001-10-23
PRIOR APPLICATION NUMBER: US 10/052,482
PRIOR FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 09/997,722
PRIOR FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: US 10/034,650
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: US 10/085,117
PRIOR FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: US 10/087,192
PRIOR FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 10/322,281
PRIOR FILING DATE: 2002-12-17
PRIOR APPLICATION NUMBER: US 10/322,696
PRIOR FILING DATE: 2002-12-17
NUMBER OF SEQ ID NOS: 1441
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1264
LENGTH: 339234
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(339234)
OTHER INFORMATION: n = A,T,C or G
US-10-669-920-1264

Query Match
Best Local Similarity 5.1%; Score 33.6; DB 6; Length 339234;
Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 149 GACAAATCTTAATGAACACGCTATGACTGTGAAAAACAATTTGTATGCGAAACCTTA 208
    |||||
Db 68076 GAAACATCAATCAATGTCACATGAGGTATTAAGTAAAGTCAATGCAATGAAATATTA 68135

QY 209 AGGATCAACAGGAATCTGTGTGCGCGTGAGATCAACGTTGCAATATTGTAAGCT 268
    |||||
Db 68136 ATCATTCAGAGAAATGCAATGCTGTAGCAATCTGTTCAATTTGAATATTAATCTTAGCTT 68195

QY 269 ATAAAAATGCTTCGAGTAGAGAAAAAAGAAATTTTGA 308
    |||||
Db 68196 ATTAATTTGAATTTTATTAAGATACCAATGATTAATTA 68235

RESULT 11
US-11-266-748A-207141/C
Sequence 207141, Application US/11266748A
Publication No. US2006013463A1
GENERAL INFORMATION:
APPLICANT: Harkin, Paul
```

```
APPLICANT: Johnston, Patrick
APPLICANT: Mulligan, Karl
TITLE OF INVENTION: Transcriptome Microarray Technology and
FILE REFERENCE: 55815-0102 (319189)
CURRENT APPLICATION NUMBER: US/11/266,748A
CURRENT FILING DATE: 2005-11-03
PRIOR APPLICATION NUMBER: EP 04105479.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105482.6
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105483.4
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105507.0
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105485.9
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105484.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: US 60/662,276
PRIOR FILING DATE: 2005-03-14
PRIOR APPLICATION NUMBER: US 60/700,293
PRIOR FILING DATE: 2005-07-18
NUMBER OF SEQ ID NOS: 483996
SOFTWARE: PatentIn version 3.3
SEQ ID NO 207141
LENGTH: 1000
TYPE: DNA
ORGANISM: Homo Sapiens
US-11-266-748A-207141
```

```
Query Match
Best Local Similarity 55.7%; Score 33.4; DB 8; Length 1000;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 109 TATCTGTAGACGAGATGTATGCAAGCCCATTAAGATAGACAAATCTTAATGAACAC 168
Db 797 TAGCCATTAGTATATGTCATATTCGAGCATTTGAATGTGGCGAGTGAAG 738

QY 169 GGCTAGACTGCAAAACATTTGTATACGAAACCTTAAGCATCAACAGAA 223
Db 737 TGTATTAAGTGAATAATATATCTGATTTTGAACCTTAGTGAATAAAGAA 683
```

```
RESULT 12
US-10-641-321-205/c
Sequence 205, Application US/10641321
GENERAL INFORMATION:
APPLICANT: Berlin, Kurt
APPLICANT: Olek, Alexander
APPLICANT: Beck, Stephan
APPLICANT: Hildmann, Thomas
APPLICANT: Lewin, Jorn
APPLICANT: Novik, Karen
TITLE OF INVENTION: Methods and Compositions for Differentiating Tissues or Cell Type
FILE REFERENCE: 47675-49
CURRENT APPLICATION NUMBER: US/10/641,321
CURRENT FILING DATE: 2003-08-12
NUMBER OF SEQ ID NOS: 221
SEQ ID NO 205
LENGTH: 4647455
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (4267840)
OTHER INFORMATION: unknown base
US-10-641-321-205
```

```
Query Match
Best Local Similarity 53.4%; Score 33.4; DB 6; Length 4647455;
```

```
Matches 70; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 176 ACTGTGAAAAACAATTTGTATCGGAAACCTTAAGGCATCAACAGAACTTGCTGTGG 235
Db 241817 ACTGTGTACAGATACCAATATCGAGCACCTTATCTGAACAAATGGCTTACTGATCA 241758

QY 236 CGTGTAGCTACACGCTTGCAATATTTGTATGATATTAATAATGCTTCCGAGTAACGAAA 295
Db 241757 GGTGTACTACGTAATAATATGCTTAAGATATAGCTATTAATTAAGCATTAAGAGGGCATG 241698

QY 296 AAAGAATTAAT 306
Db 241697 GAAAAATTTTT 241687
```

```
RESULT 13
US-11-266-748A-411280/c
Sequence 411280, Application US/11266748A
GENERAL INFORMATION:
APPLICANT: Johnston, Patrick
APPLICANT: Mulligan, Karl
TITLE OF INVENTION: Transcriptome Microarray Technology and
FILE REFERENCE: 55815-0102 (319189)
CURRENT APPLICATION NUMBER: US/11/266,748A
CURRENT FILING DATE: 2005-11-03
PRIOR APPLICATION NUMBER: EP 04105479.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105482.6
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105483.4
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105507.0
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105485.9
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: EP 04105484.2
PRIOR FILING DATE: 2004-11-03
PRIOR APPLICATION NUMBER: US 60/662,276
PRIOR FILING DATE: 2005-03-14
PRIOR APPLICATION NUMBER: US 60/700,293
PRIOR FILING DATE: 2005-07-18
NUMBER OF SEQ ID NOS: 483996
SOFTWARE: PatentIn version 3.3
SEQ ID NO 411280
LENGTH: 397
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (90)..(90)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (126)..(126)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (220)..(220)
OTHER INFORMATION: n is a, c, g, or t
US-11-266-748A-411280
```

```
Query Match
Best Local Similarity 52.6%; Score 33.2; DB 8; Length 397;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
QY 329 GTTCGTAAACAGATACAGATGAGAAACGCTTGCTTAAGACCTTGGAGATGTCAT 388
Db 341 GTCTTAATAACATTCACAGATGACCAATATGATATAGCAAGTATTTGTTCTCAT 282

QY 389 CGTTTCTCTTATGCTAATATCTGACGAGAAATGTTATTAAGAAAGTCTAGTAATCTT 448
```

Mon Sep 18 16:01:18 2006

US-10-815-774-5.rnpbn

Db 281 CAGTTTCACTTAAGTATGCGATTAAAGCCATATTATGCTTCCAGTGTATGAT 222  
QY 449 ACCGTATGACACAA 463  
Db 221 GNAATTAGTTAA 207

RESULT 14  
US-11-266-748A-206397/C  
; Sequence 206397, Application US/11266748A  
; Publication No. US20060134663A1  
; GENERAL INFORMATION:  
; APPLICANT: Harkins, Paul  
; APPLICANT: Johnston, Patrick  
; APPLICANT: Mulligan, Karl  
; TITLE OF INVENTION: Transcription Microarray Technology and  
; TITLE OF INVENTION: Methods of Using the Same  
; FILE REFERENCE: 55815-0102 (319189)  
; CURRENT APPLICATION NUMBER: US/11/266,748A  
; PRIOR APPLICATION NUMBER: EP 04105479.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105482.6  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105483.4  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105507.0  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105485.9  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: EP 04105484.2  
; PRIOR FILING DATE: 2004-11-03  
; PRIOR APPLICATION NUMBER: US 60/662,276  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/700,293  
; PRIOR FILING DATE: 2005-07-18  
; NUMBER OF SEQ ID NOS: 483996  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 206397  
; LENGTH: 1000  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-11-266-748A-206397

Query Match 5.9%; Score 33; DB 8; Length 1000;  
Best Local Similarity 50.3%; Pred. No. 3.2;  
Matches 81; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 140 ATAGATGAGACAAATCTTAATGACACGCGTATGACTGTGAAAAACAATTGTATACG 199  
Db 623 ATAGATGATGATCATGTAATGAAAGACACTGAAAAACATTAATAAAAAAATGTAAAGT 564  
QY 200 GAAACCTTAAGGATCAACGAGAACTGTGTGGCTGAGCTACCAAGTTGCATAT 259  
Db 563 CAAACCAAAACCTTAACATTAACACGCACTGACACATAGCAAGTCTATCC 504  
QY 260 TGGTAACTATTAATAATGCTTCCGAGTAAAGAAAAAGA 300  
Db 503 TGGACACGAAAGTCCGATTTCTTTAAATGCCAAAAA 463

RESULT 15  
US-10-953-349-36486  
; Sequence 36486, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30

; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 36486  
; LENGTH: 1467  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
US-10-953-349-36486

Query Match 5.9%; Score 33; DB 6; Length 1467;  
Best Local Similarity 57.9%; Pred. No. 3.9; Indels 1; Gaps 1;  
Matches 77; Conservative 0; Mismatches 55

QY 372 CTCTGGGATTCGATCCGTTCTCTATGCTAATCTGACGAAATGTTTATACAG 431  
Db 59 CTACTATGATTTCTGCACTGCAATCTGATGAACTACTGCGCTTACTATGACAG 118  
QY 432 AAGTCTAGTAAATCTTACCTGTATGA-CACAAATCTGATCAATCAACTGTGACTGA 490  
Db 119 CAACCTGAGTTGGATTCATATGATCAACAGACTCAACAAATATGTCCTTGTATGA 178  
QY 491 CTAAATTTGACT 503  
Db 179 CCAGAGCAAGACT 191

Search completed: September 16, 2006, 18:38:33  
Job time : 203 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2006, 14:56:20 ; Search time 3896 Seconds  
(without alignments)  
7965.921 Million cell updates/sec

Title: US-10-815-774-5  
Perfect score: 555  
Sequence: 1 ttgcgcagcgtatccctcga.....atgtatccagctggtgattt 555

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 48236798 seqs, 2795965780 residues

Total number of hits satisfying chosen parameters: 96473536

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: gb\_est1:  
2: gb\_est3:  
3: gb\_est4:  
4: gb\_est5:  
5: gb\_est6:  
6: gb\_est7:  
7: gb\_est8:  
8: gb\_est9:  
9: gb\_est10:  
10: gb\_est11:  
11: gb\_est12:  
12: gb\_est13:  
13: gb\_est14:  
14: gb\_est15:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.2	7.4	706	12	CE750748
2	40.8	7.4	593	4	CBS584627
3	40.6	7.4	606	14	FR0009895
4	40.6	7.3	876	9	DN089815
5	40.4	7.3	233	1	AA767965
6	40	7.2	600	14	DE045970
7	40	7.2	825	14	DE260973
8	39.4	7.1	438	2	BI379434
9	39.4	7.1	498	2	BI379434
10	39.4	7.1	567	4	BM835038
11	39	7.0	532	2	BI893381
12	39	7.0	570	7	AM937387
13	39	7.0	570	7	AM937408
14	38.6	7.0	374	5	CJ349523
15	38.4	6.9	2553	6	AK086154
16	38	6.8	342	5	CJ391019
17	38	6.8	539	4	BM834458
18	38	6.8	782	5	CK861052
19	37.8	6.8	716	4	BM784868

20	37.6	6.8	729	3	BP020887
21	37.6	6.8	858	7	BF697486
22	37.6	6.8	932	14	CR240908
23	37.6	6.8	958	14	CR276499
24	37.4	6.7	401	8	CN757303
25	37.4	6.7	401	8	CN757304
26	37.4	6.7	489	3	BM885104
27	37.4	6.7	531	2	BM092866
28	37.4	6.7	899	8	CV832934
29	37.4	6.7	961	12	CU091135
30	37.2	6.7	728	14	AG164728
31	37.2	6.7	784	4	BM953380
32	37	6.7	496	12	CE347872
33	37	6.7	561	9	DN254639
34	37	6.7	689	9	DN265436
35	37	6.7	709	9	DN267374
36	37	6.7	713	9	DN267044
37	36.8	6.6	550	11	AQ982897
38	36.8	6.6	692	9	CX577073
39	36.8	6.6	1052	14	DU780626
40	36.6	6.6	555	2	BG683527
41	36.6	6.6	820	13	CM938378
42	36.6	6.6	883	7	BB898984
43	36.4	6.6	422	14	DR7K7S
44	36.4	6.6	433	12	BZ685656
45	36.4	6.6	491	11	AQ143297

## ALIGNMENTS

RESULT 1  
LOCUS CE750748/c 706 bp DNA linear GSS 30-SEP-2003  
DEFINITION tigr-gss-dog-17000369607113 Dog library Canis familiaris genomic.  
ACCESSION CE750748  
VERSION CE750748.1 GI:37091365  
KEYWORDS GSS  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.  
REFERENCE 1 (bases 1 to 706)  
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.  
TITLE The dog genome: survey sequencing and comparative analysis  
JOURNAL PUBMED  
PUBMED 14512627  
COMMENT Contact: Kirkness EF  
The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirkness@tigr.org  
Class: shotgun.  
FEATURES  
source  
1..706  
Location/Qualifiers  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"  
ORIGIN  
Query Match 7.4%; Score 41.2; DB 12; Length 706;  
Best Local Similarity 55.6%; Pred. No. 2.2;  
Matches 79; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 207 AAGGATCAACGAACTTGTGCGTGAGCTACCACTTCAATATTGTAAG 266  
 DB 144 AGGATGACAAATGCTGTGTGATGTGAGAAAGAAATCCATGTGCACTGTGTGG 85  
 QY 267 CTATAAAATGCTCCGAGTAAAGGAAATATTGATCTTCACTATTTCAG 326  
 DB 84 ATTGAAATTTGGCGACTACTATGAAATATGATGAGATCTCACAATTTAAAA 25  
 QY 327 CGGTCTGTACAGATACCA 348  
 DB 24 CAGAACTACCAAGATCCAGCA 3

RESULT 2  
 CB584627  
 LOCUS  
 DEFINITION AMGNNUC:NRHY4-00187-A11-A W Rat hypothalamus (10464) Rattus  
 norvegicus cDNA clone nrhy4-00187-a11 5', mRNA sequence.  
 ACCESSION CB584627  
 VERSION  
 KEYWORDS EST.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.  
 1 (bases 1 to 593)  
 Amgen EST Program.  
 Amgen Rat EST Program  
 Unpublished (2003)  
 CONTACT: Dan Fitzpatrick  
 Amgen, Inc  
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00187 row: a column: 11.  
 Location/Qualifiers  
 1..593  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrhy4-00187-a11"  
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; W Rat  
 hypothalamus adult female Wistar rat avg. insert size 2.3  
 kb fraction 6 and 7"

ORIGIN  
 Query Match 7.4%; Score 40.8; DB 4; Length 593;  
 Best Local Similarity 55.7%; Pred. No. 2.8; Indels 0; Gaps 0;  
 Matches 78; Conservative 0; Mismatches

QY 289 ACGGAAAAAGATATTGATCTTCACTATTTCAGCGGTCTGTACAGATACCA 348  
 DB 280 AATATTAACAAATCAATCTTCAATGTTAAGAAAGTTCTGGCAACCAACCA 339  
 QY 349 TGGAGAAAGCTTGGCTTAACACCTTGGCGATCTGCATCGCTTCTTAACTAT 408  
 DB 340 CCTAAAAAGACTGATTAAGAGCTGTCTGAATATGATGCTCTTCTTAATGGCAAT 399  
 QY 409 ACTGCAGGAAATGTTATTA 428  
 DB 400 ACTTAATTAATGCTTAATTA 419

RESULT 3  
 FR0009895/c  
 LOCUS FR0009895 606 bp DNA linear GSS 25-FEB-2004  
 DEFINITION F. rubripes GSS sequence, clone 032C09AD10, genomic survey sequence.  
 ACCESSION AL001177  
 VERSION AL001177.1 GI:2439029  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Takifugu rubripes (Pugu rubripes)

ORGANISM Takifugu rubripes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Acinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodontidae; Takifugu.  
 1 (bases 1 to 606)  
 Elgar, G., Clark, M., Smith, S., Warner, S., Edwards, Y., J.,  
 Bouchireb, N., Cottage, A., Yeo, G.S., Umanita, Y., Williams, G., and  
 Brenner, S.  
 Generation and analysis of 25 Mb of genomic DNA from the pufferfish  
 Fugu rubripes by sequence scanning  
 Genome Res. 9 (10), 960-971 (1999)  
 10523524  
 2 (bases 1 to 606)  
 Elgar, G., Clark, M., Smith, S., Warner, S., Warner, S., Umanita, Y.,  
 Williams, G., and Brenner, S.  
 Direct Submission  
 Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource  
 Centre Hinxton, Cambridge, CB10 1SB. Email: biolhelp@hmp.mrc.ac.uk  
 Vector: plasmidscript II KS  
 V-type: plasmid  
 PRIMER: KS  
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic  
 sequence.  
 Location/Qualifiers  
 1..606  
 /organism="Takifugu rubripes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:31033"  
 /clone="032C09AD10"  
 /clone\_lib="cosmid 032C09"

ORIGIN  
 Query Match 7.4%; Score 40.8; DB 14; Length 606;  
 Best Local Similarity 50.3%; Pred. No. 2.8; Indels 0; Gaps 0;  
 Matches 93; Conservative 0; Mismatches

QY 183 AAAAAATTTGTATCGGAAACCTTAAGCATCAAGAACTTGCTGTGGCGTGAG 242  
 DB 553 AGATATATTGTATGTGCTGATGACGAAAGTATCCATCCCACTTTNAGCGTTTGGT 494  
 QY 243 CTACACGTTGCAATATTGTAAGCTATTAATAATCTTCCGAGTAAGGAAAAAGAT 302  
 DB 493 AAGATATGCTGCAAAAGTGAAGCTAAATATATATCTATATGCGCAAGTNGAT 434  
 QY 303 TATTGATCTTCACTATTTTCAAGCGCTCTGTACAGATACAGATGAGAAAGCTTG 362  
 DB 433 TTATCTCCATCTTATTTGAAATTTCTGTAGAGATAGATGTCATTAAGCGCGG 374  
 QY 363 CGTTA 367  
 DB 373 CTGCA 369

RESULT 4  
 DN089815/c  
 LOCUS DN089815 876 bp mRNA linear EST 14-FEB-2005  
 DEFINITION JGI CABE3015.fwd NIH XGC tropoval xenopus tropicalis cDNA clone  
 IMAGE:7821900 5', mRNA sequence.  
 ACCESSION DN089815  
 VERSION DN089815.1 GI:59756919  
 KEYWORDS EST.  
 SOURCE Xenopus tropicalis (western clawed frog)  
 ORGANISM Xenopus tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Xenopus; Silurana.  
 1 (bases 1 to 876)  
 Richardson, P., Lucas, S., Rokhsar, D., Dettler, J.C., Ng, D.C.,  
 Brokstein, P., and Lindquist, E.A.  
 DOE Joint Genome Institute Xenopus tropicalis EST project  
 Unpublished (2004)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

## COMMENT

Other\_ESTs: JGI CABE3015.rev  
Contact: Lindquist, E.A., Richardson, P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: cdna@jgi-psf.org  
Tissue Procurement: Robert M. Grainger  
CDNA Library Preparation: Bruce Blumberg Laboratory, University of California, Irvine  
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
<http://image.llnl.gov>  
Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix 'fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.  
Plate: CABE 0029 row: n column: 10  
High quality sequence stop: 846.  
Location/Qualifiers

## FEATURES

source

```
1. 876
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clone="IMAGE:7821900"
/sex="Female"
/tissue_type="Ovary"
/dev_stage="Adult"
/lab_host="ElectroMAX DH10B TI Phage Resistant cells"
/clone_1lb="NIH XCC tropoval"
/notes="Vector: PCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming
(5'-ACTAGTCGGCCGCCCTCAGCCTTCAGCTTTTCTTTTCTTTT-3') and
Stratascript reverse transcriptase. After ligation of
EcoRI adapters (5'-AATTCGGCAGAG-3') followed by kinasing
adapters and by XhoI digestion, the cDNA was size selected
by chromatography on Sepharose CL-2B columns and fractions
containing cDNAs larger than 1000 bp were ligated into
EcoRI/XhoI-digested PCS107. Reference for library
construction: Current Genomics 4, 635-644. Library
constructed by Michelle Tabb and Bruce Blumberg (Dept of
Developmental and Cell Biology, University of California,
Irvine)."
```

## ORIGIN

Query Match 7.3%; Score 40.6; DB 9; Length 876;  
Best Local Similarity 54.3%; Pred. No. 3.5;  
Matches 82; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
QY 128 ATGCAAGACCCATTAAGATGAGACAAATCTTAATGAACAACGGCTATGCTGAAAAAC 187
DB 307 ATTAACAACCCATTAATGAACAACAGTATGATGAATACCTTTGCTTGAAAAAT 248
QY 188 AATTTGTATACGAAACCTTAAGGATCAACAGAACTTGCTGTGAGGAGTAC 247
DB 247 ATACTTTTACTGCCATTTTAAGTACAAACATCCCTTGTTGTGAGAGGGCTTTATT 188
QY 248 ACGTTGCATATTTGTAGCTATTAATAATGC 278
DB 187 AATGAGACAAATGTGAGGCTGTGCTGTGC 157
```

RESULT 5  
AA767965 233 bp mRNA linear EST 07-FEB-1998  
LOCUS n96612.s1 NCI CGAP\_Prl6 Homo sapiens cDNA clone IMAGE:1237630,  
DEFINITION mRNA sequence.  
ACCESSION AA767965  
VERSION AA767965.1 GI:2818980  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Homo.  
REFERENCE  
1 (bases 1 to 233)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: W. Maeston Linehan, M.D., Rodrigo F. Chuquai,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 332 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amerham.  
Location/Qualifiers

## FEATURES

source

```
1. 233
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1237630"
/sex="male"
/tissue_type="tumor"
/lab_host="DH10B"
/clone_1lb="NCI CGAP_Prl6"
/notes="Organ: prostate; Vector: PAMPI0; mRNA made from  
invasive prostate tumor cells, cDNA made by oligo-dT  
priming. Non-directionally cloned. Size-selected on  
agarose gel, average insert size 600 bp. Library made by  
D. Krizman, NIH."
```

## ORIGIN

Query Match 7.3%; Score 40.4; DB 1; Length 233;  
Best Local Similarity 49.5%; Pred. No. 2.8;  
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```
QY 18 TGATGTAGCTACATTAAATTTCTTAATCAATCAATAAAGATCAGTCTTGCGTACTC 77
DB 21 TGGTTTGTAAACATACATATATACAGATTCACATATGATTTACTTGTAGGCTTC 80
QY 78 TAGCGCGTCTCACCATGATCATTCATGATATCTGTAGCGAGTGTATGCAAGAC 137
DB 81 TTTTGTCCAAACCATGCGGTTCCATACAGAGCCTCTGACACATGTAGCTGATGAC 140
QY 138 CCATAAGATGAGACAAATCTTAATGACAAAGCGCTATGACTGTGAACAAATTTGTATA 197
DB 141 AGATTAATATGCTAGTCCAAATTAAGGTATCTGTAGTATTAATGAAGAGTGATT 200
QY 198 CGGAACCTTAAGGATCAACAGAACTTG 227
DB 201 TCCAAACAGTATTCAAAAAAGATCG 230
```

RESULT 6  
DE045970/c 600 bp DNA linear GSS 25-MAY-2005  
LOCUS Oryzias latipes DNA, clone: olal-011D04.F, genomic survey sequence.  
DEFINITION DE045970  
ACCESSION DE045970  
VERSION DE045970.1 GI:62557514  
KEYWORDS GSS.  
SOURCE Oryzias latipes (Japanese medaka)  
ORGANISM Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.  
REFERENCE 1

**AUTHORS** Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.  
**TITLE** BAC end sequences of Olai Oryzias latipes Library  
**JOURNAL** Published Only in Database (2005)  
**REFERENCE** 2 (bases 1 to 600)  
**AUTHORS** Fujiyama, A.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenhiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: atfujiam@sc.riken.jp, URL: http://stc.gsc.riken.jp/, Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)  
**COMMENT** This work was done in collaboration with Takeda, H. (1), Naruse, K. (2) and Narita, T. (3)  
 (1) Department of Biological Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: htakeda.s.u-tokyo.ac.jp  
 (2) Department of Biological Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: naruse.s.u-tokyo.ac.jp  
 (3) Department of Biological Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: tanarita.s.u-tokyo.ac.jp  
**PRIMERS**  
**Sequencing** : Forward  
**LIBRARY**  
 Vector : pKS145  
 R.Site 1 : SacI  
 L.Site 2 : SacI  
**Location/Qualifiers**  
 1. 600  
 /organism="Oryzias latipes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:8090"  
 /clone="olai-011D04.F"  
 /sex="male"  
 /cell\_type="whole body"  
 /clone\_lib="BAC end sequences of Olai Oryzias latipes library"

**FEATURES**  
**source**

**ORIGIN**  
 Query Match 7.2%; Score 40; DB 14; Length 600;  
 Best Local Similarity 55.9%; Pred. No. 4.7;  
 Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

36 TTCTTATTCATCAATAAAGATCAGTCTTGCGGTACCTCTACGGCGTCTCACCATG 95  
 |||||  
 402 TTCTTATTTTAAACAAATTAACAAGCCTGCTAGTTGGGCTGCAAGACATGCTCACA 343  
 |||||  
 96 CATCAATTCAGATATCTGTGAACGAGTGTATGCAAGAGCCCATAGATGACAAAT 155  
 |||||  
 342 AATGCAATTCAGATTGCAATTAAAGAGATTCTGGCTGTTGCTACATATGATCAAAA 283  
 |||||  
 156 CTTAATGACACAGCGC 171  
 |||||  
 282 ACTAAAAACAATGGC 267

**RESULT** 7  
 DE260973/c DE260973 825 bp DNA linear GSS 22-SEP-2005  
**LOCUS** Oryzias latipes DNA, clone: olai-148A13.F, genomic survey sequence.  
**DEFINITION** DE260973  
**ACCESSION** DE260973.1 GI:76077885  
**VERSION**

**KEYWORDS** GSS.  
**SOURCE** Oryzias latipes (Japanese medaka)  
**ORGANISM** Oryzias latipes  
**REFERENCE** 1  
**AUTHORS** Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.  
**TITLE** BAC end sequences of Olai Oryzias latipes Library  
**JOURNAL** Published Only in Database (2005)  
**REFERENCE** 2 (bases 1 to 825)  
**AUTHORS** Fujiyama, A.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (16-SEP-2005) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenhiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: atfujiam@sc.riken.jp, URL: http://stc.gsc.riken.jp/, Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)  
**COMMENT** This work was done in collaboration with Takeda, H. (1), Naruse, K. (2) and Narita, T. (3)  
 (1) Department of Biological Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: htakeda.s.u-tokyo.ac.jp  
 (2) Department of Biological Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: naruse.s.u-tokyo.ac.jp  
 (3) Department of Biological Science, University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: tanarita.s.u-tokyo.ac.jp  
**PRIMERS**  
**Sequencing** : Forward  
**LIBRARY**  
 Vector : pKS145  
 R.Site 1 : SacI  
 L.Site 2 : SacI  
**Location/Qualifiers**  
 1. 825  
 /organism="Oryzias latipes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:8090"  
 /clone="olai-148A13.F"  
 /sex="male"  
 /cell\_type="whole body"  
 /clone\_lib="BAC end sequences of Olai Oryzias latipes library"

**FEATURES**  
**source**

**ORIGIN**  
 Query Match 7.2%; Score 40; DB 14; Length 825;  
 Best Local Similarity 55.9%; Pred. No. 5.1;  
 Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

36 TTCTTATTCATCAATAAAGATCAGTCTTGCGGTACCTCTACGGCGTCTCACCATG 95  
 |||||  
 397 TTCTTATTTTAAACAAATTAACAAGCCTGCTAGTTGGGCTGCAAGACATGCTCACA 338  
 |||||  
 96 CATCAATTCAGATATCTGTGAACGAGTGTATGCAAGAGCCCATAGATGACAAAT 155  
 |||||  
 337 AATGCAATTCAGATTGCAATTAAAGAGATTCTGGCTGTTGCTACATATGATCAAAA 278  
 |||||  
 156 CTTAATGACACAGCGC 171  
 |||||  
 277 ACTAAAAACAATGGC 262

RESULT 8  
 BI379434/c 438 bp mRNA linear EST 26-ANG-2003  
 LOCUS BFLG1\_001065 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG  
 DEFINITION or MEMGP498) Branchiostoma floridae cDNA clone MEMGP498E1916 5',  
 mRNA sequence.  
 BI379434  
 VERSION BI379434.1 GI:30914584  
 ACCESSION  
 KEYWORDS EST.  
 SOURCE Branchiostoma floridae (Florida lancelet)  
 ORGANISM Branchiostoma floridae  
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 Branchiostoma.  
 REFERENCE 1 (bases 1 to 438)  
 AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Pousetka, A.J.,  
 Herwig, R., Vingron, M. and Lehrach, H.  
 TITLE New evidence for genome-wide duplications at the origin of  
 vertebrates using an amphioxus gene set and completed animal  
 genomes  
 JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)  
 PUBMED 12799346  
 COMMENT Contact: Panopoulou G  
 Laboratory 145, dept. Lehrach  
 Max-Planck-Institut fuer Molekulare Genetik  
 Ihnestr. 63-73, D-14195 Berlin, Germany  
 Tel: +49 30 8413 1235  
 Fax: +49 30 8413 1128  
 Email: panopoulou@molgen.mpg.de  
 The library was characterised by oligonucleotide fingerprinting  
 (ONFP) to reduce sequencing redundancy. According to the ONFP  
 procedure, clones giving the same hybridisation pattern with a  
 battery of 200 8mer oligonucleotides are grouped into clusters. One  
 clone per cluster is selected for sequencing. The size of each  
 cluster is an indicator of the frequency of a transcript in the  
 analysed library. The cluster size as well as the coordinates of the  
 rest of the clones assigned to the same fingerprint cluster as the  
 clone from which the above EST is generated is available at the  
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>  
 Clones and filters are distributed via the Resource Centre/Primary  
 Database of the German Genome Project (<http://www.rzpd.de>)  
 PCR PRIMERS  
 FORWARD: 5' CCCAGGCTTTACCTTATGCTCCGGCTCG 3' (M13RSP)  
 BACKWARD: 5' GCTATTACGCCAGCTGGCGAAGGGGATGTC 3' (M13RSP)  
 Insert Length: 1200 Std Error: 0.00  
 Seq primer: 5'-CCGCTCCGGAATTCGCGGT-3' pSPORT3/86  
 High quality sequence stop: 438.  
 Location/Qualifiers  
 1..438  
 /organism="Branchiostoma floridae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7739"  
 /clone="MEMGP498E1916"  
 /tissue\_type="whole embryo"  
 /dev\_stage="5-6 hrs (gastrula stage)"  
 /lab\_host="E.coli, XL1 blue"  
 /clone\_lib="Amphioxus 5-6 hrs cDNA library (Name  
 convention: BFLG or MEMGP498)"  
 /note="Vector: pSPORT1; Site 1: SalI, KpnI, SmaI (5');  
 Site 2: NotI, BamHI, HindIII (3'); OligodT primed and  
 directionally cloned in pSPORT1 vector using a NotI  
 (5'-pGACTAGTTCTGATCGCGAGCGCGGCC (T15-3' and a SalI 5'-  
 TCGACCCAGCGCTCG-3' adapters (Gibco BRL)."

## ORIGIN

Query Match 7.1%; Score 39.4; DB 2; Length 438;  
 Best Local Similarity 57.9%; Pred. No. 6.5;  
 Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 394 TCCCTTATGCTTAATCTGACGAGAAATGTTTATTACGAGAGCTTATCTTACCTG 453  
 DB 267 TACTCTTAAGCTACACAGCAAGAACGTAATCAAGAGCTTACAAATCTTACCTA 208

QY 454 TATGACAAATCTGATCAATACCACTGTGTACTGACTAATTTTCACGCTTCGGA 513  
 DB 207 TGGCATTAATTAATAAATAATTTGGAAGTGTCTATTAATAATTAAGTGTGACTTAA 148  
 QY 514 T 514  
 DB 147 T 147

RESULT 9  
 BW835568 498 bp mRNA linear EST 23-OCT-2005  
 LOCUS BW835568 Amphioxus Branchiostoma floridae unpublished cDNA library,  
 DEFINITION larva whole animal Branchiostoma floridae cDNA clone bflv055g01 3',  
 mRNA sequence.  
 BW835568  
 VERSION BW835568.1 GI:66443784  
 ACCESSION  
 KEYWORDS EST.  
 SOURCE Branchiostoma floridae (Florida lancelet)  
 ORGANISM Branchiostoma floridae  
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 Branchiostoma.  
 REFERENCE 1 (bases 1 to 498)  
 AUTHORS Yu, J., Holland, L.Z., Shin, I.T., Kohara, Y., Satou, Y. and Satoh, N.  
 TITLE Expressed genes in Branchiostoma floridae  
 JOURNAL Unpublished (2005)  
 COMMENT Contact: Tadao Shin-i  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp  
 If you want to have a cDNA clone for this EST or if you have any  
 questions, please send an e-mail to Nori Satoh  
 (satoh@nig.ac.jp) and its cc to Yutaka Satou  
 (yutaka@nig.ac.jp).  
 Location/Qualifiers  
 1..498  
 /organism="Branchiostoma floridae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7739"  
 /clone="bflv055g01"  
 /tissue\_type="larva"  
 /dev\_stage="larva"  
 /clone\_lib="Amphioxus Branchiostoma floridae unpublished  
 cDNA library, larva whole animal"

## ORIGIN

Query Match 7.1%; Score 39.4; DB 4; Length 498;  
 Best Local Similarity 57.9%; Pred. No. 6.7;  
 Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 394 TCCCTTATGCTTAATCTGACGAGAAATGTTTATTACGAGAGCTTATCTTACCTG 453  
 DB 112 TACTCTTAAGCTACACAGCAAGAACGTAATCAAGCGCTTACAAATATTAACCTA 171  
 QY 454 TATGACAAATCTGATCAATACCAACTGTGTACTGACTAATTTTCACGCTTCGGA 513  
 DB 172 TGGTATTAATTAATAAATAATTTGGAAGTGTCTATTAATAATTAAGTGTGACTTAA 231  
 QY 514 T 514  
 DB 232 T 232

## RESULT 10

QY 567 bp mRNA linear EST 22-OCT-2005  
 LOCUS BW835038 Amphioxus Branchiostoma floridae unpublished cDNA library,  
 DEFINITION larva whole animal Branchiostoma floridae cDNA clone bflv053o1 3',  
 mRNA sequence.

ACCESSION BM835038  
 VERSION BM835038.1 GI:66443254  
 KEYWORDS EST.  
 SOURCE Branchiostoma floridae (Florida lancelet)  
 ORGANISM Branchiostoma floridae  
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.

REFERENCE 1 (bases 1 to 567)  
 AUTHORS Yu, J., Holland, L. Z., Shin, I. T., Kohara, Y., Satou, Y. and Satoh, N.  
 TITLE Expressed genes in Branchiostoma floridae  
 JOURNAL Unpublished (2005)  
 COMMENT Contact: Tadashi Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp  
 If you want to have a cDNA clone for this EST or if you have any questions, please send an e-mail to Nori Satoh (satoh@sci.riken.go.jp) and its cc to Yutaka Satou (yutaka@sci.riken.go.jp).

FEATURES  
 source  
 1..567  
 /organism="Branchiostoma floridae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7739"  
 /clone="B1V053019"  
 /tissue\_type="whole animal"  
 /dev\_stage="larva"  
 /clone\_lib="Amphioxus Branchiostoma floridae unpublished cDNA library, larva whole animal"

ORIGIN  
 Query Match 7.1%; Score 39.4; DB 4; Length 567;  
 Best Local Similarity 57.9%; Pred. No. 6.9;  
 Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 394 TCTCTTATGCTAATCTGACGAAGTGTATTACGAGTCTGATTAATCTTACCG 453  
 |||||  
 DB 114 TACTCTTACCTCAACAGAGAGAGTAAATCAAGCTCTTCAAAAATATTAACCTA 173  
 454 TATGACACATCTGATCATACCACTGCTACTGCTAAATTTTCACTGCTTCCGGA 513  
 DB 174 TGTATTAATTAATAAATTTGGAAGTGTGCTATTAATAATTAAGTGTATCTATA 233

QY 514 T 514  
 DB 234 T 234

RESULT 11  
 LOCUS BI893381 532 bp mRNA linear EST 08-JUL-2004  
 DEFINITION sai65d11.y1 Gm-cl068 glycine max cDNA clone GENOME SYSTEMS CLONE  
 ACCESSION BI893381  
 VERSION BI893381.1 GI:16105641  
 KEYWORDS EST.  
 SOURCE Glycine max (soybean)  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eucosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 532)  
 AUTHORS Khanna, R., Keim, P., Vodkin, L., Eipelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, J., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.  
 TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)  
 COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 425.

FEATURES  
 source  
 1..532  
 /organism="Glycine max"  
 /mol\_type="mRNA"  
 /cultivar="Williams 82"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl068-3621"  
 /tissue\_type="leaf, drought stressed, 1 month old plants, greenhouse grown"  
 /lab\_host="DH10B"  
 /clone\_lib="Gm-cl068"  
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI. The cDNA library was constructed from mRNA isolated from drought stressed leaf tissue of the cultivar Williams 82. The month old greenhouse grown plants were deprived of water for 3 days prior to harvesting the stressed leaf tissue. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dt) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

ORIGIN  
 Query Match 7.0%; Score 39; DB 2; Length 532;  
 Best Local Similarity 68.4%; Pred. No. 8.9;  
 Matches 54; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 34 AATCTTATTCATCAATAAAGATCGCTTGGGACCTCTACGGCGTCCACCA 93  
 |||||  
 DB 5 AATCTTTGTCCAAACAAAGATGATGATGCTGTTCTTACGAGATGTCACCA 64

QY 94 TGCATCATTCAGATATC 112  
 |||||  
 DB 65 TGAATCATCTTCAGTTACC 83

RESULT 12  
 LOCUS AW937387/c 570 bp mRNA linear EST 30-MAY-2000  
 DEFINITION OV3-DT0043-180100-061-c07 DT0043 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW937387  
 VERSION AW937387.1 GI:8112810  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 570)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Bruneau, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, U. F., de Souza, S. J. and Simpson, A. J.

**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**PUBMED** 10737800

**COMMENT** Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=kt2-QV3-DT0043-180  
100-061-c07kt3=2000-01-18kt4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 569.

**FEATURES**  
source  
1..570  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="DT0043"  
/note="Organ: denis drash; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

**ORIGIN**  
Query Match 7.0%; Score 39; DB 7; Length 570;  
Best Local Similarity 52.8%; Pred. No. 9;  
Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

182 AAAACAATTGTATACGGAACCTTAAGGATCAACAGAACTGCTGTGGCGTGA 241  
|||||  
480 AAAAGAGTTTATAGGAATACTGTTACTTGTATAGAACTTATGATTTAA 421  
242 GCTACACCTGCATATTTGTAAGCTATTAATAATCTCCGAGTACGAAAAAGAA 301  
420 GCTACCTGTATATAGGGGTAGGCTAGAGTAGCCCAAGGAAAAAGGTAATTGG 361  
QY 302 TTATTGATCCTTCACATATTTTCAAGCGGCTCTGTAAAG 340  
DB 360 CTAAAGAAATTCACAAATTTAAGTTGCTTTAAAAAG 322

**RESULT 13** AM937408 570 bp mRNA linear EST 30-MAY-2000  
LOCUS AM937408.1-190100-061-c07 DT0043 Homo sapiens cDNA, mRNA sequence.  
DEFINITION QV3-DT0043-190100-061-c07 DT0043 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM937408  
VERSION AM937408.1 GI:8112831  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo  
1 (bases 1 to 570)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**PUBMED** 10737800

**COMMENT** Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=kt2-QV3-DT0043-190  
100-061-c07kt3=2000-01-19kt4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 569.

**FEATURES**  
source  
1..570  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="DT0043"  
/note="Organ: denis drash; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

**ORIGIN**  
Query Match 7.0%; Score 39; DB 7; Length 570;  
Best Local Similarity 52.8%; Pred. No. 9;  
Matches 84; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

182 AAAACAATTGTATACGGAACCTTAAGGATCAACAGAACTGCTGTGGCGTGA 241  
|||||  
480 AAAAGAGTTTATAGGAATACTGTTACTTGTATAGAACTTATGATTTAA 421  
242 GCTACACCTGCATATTTGTAAGCTATTAATAATCTCCGAGTACGAAAAAGAA 301  
420 GCTACCTGTATATAGGGGTAGGCTAGAGTAGCCCAAGGAAAAAGGTAATTGG 361  
QY 302 TTATTGATCCTTCACATATTTTCAAGCGGCTCTGTAAAG 340  
DB 360 CTAAAGAAATTCACAAATTTAAGTTGCTTTAAAAAG 322

**RESULT 14** CJ349523 374 bp mRNA linear EST 15-JUN-2005  
LOCUS CJ349523.3 374 bp mRNA linear EST 15-JUN-2005  
DEFINITION CJ349523 Molgula tectiformis unpublished cDNA library, cleaving embryo Molgula tectiformis cDNA clone mcl013K20 5', mRNA sequence.  
ACCESSION CJ349523  
VERSION CJ349523.1 GI:67789671  
KEYWORDS EST.  
SOURCE Molgula tectiformis  
ORGANISM Molgula tectiformis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Stolidobranchia; Molgulidae; Molgula.  
1 (bases 1 to 374)  
Gyoja, F., Satou, Y. and Satoh, N.  
Expressed genes in Molgula tectiformis  
Unpublished (2005)  
Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp  
When you want to obtain this EST clone, please send an e-mail to Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to yutaka

FEATURES  
source  
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).  
Location/Qualifiers  
1.374  
/organism="Molgula tectiformis"  
/mol\_type="mRNA"  
/db\_xref="taxon:30286"  
/clone="mtc1013k20"  
/issue\_type="whole animal"  
/dev\_stage="cleaving embryo"  
/clone\_lib="Molgula tectiformis unpublished cDNA library,  
cleaving embryo"

ORIGIN  
Query Match 7.0%; Score 38.6; DB 5; Length 374;  
Best Local Similarity 50.8%; Pred. No. 11;  
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 272 AAAATGCTTCGGAGTACGGAAGAAATTTGATCTCTTCACTATTTTCAAGCGCTC 331  
DB 352 AGAGTCTTTTCCAAAGTAAAGTTAAATGAATCCATTTGCTTACGAAATGAACTTTTC 293  
QY 332 CTGTAAAGATACAGCATGAGAAAGCTTGGCTTACACCTTGGCGATCTGACCTCG 391  
DB 292 TGTTTAAGATGAAGAAATTTGATCATATATTTGAAGTAAACAGTACGGAAGCTGACCTT 233  
QY 392 TTTCCTCTTATGCTAATATCTGAGAAATGTTTATTAACAGAGTCTTATTTTAC 451  
DB 232 TGTATATGGGCTTTTACACAGCTGCTTCTCAATGAGTAAATCTTTTAAAGTGTTC 173  
QY 452 T 452  
DB 172 T 172

RESULT 15  
AK086154 2553 bp mRNA linear HTC 02-SEP-2005  
LOCUS Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched  
DEFINITION library, clone:D930008124 product:unclassifiable, full insert  
sequence.  
ACCESSION AK086154  
VERSION AK086154.1 GI:26103259  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
PUBMED 10349636  
REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes  
PUBMED Genome Res. 10 (10), 1617-1630 (2000)  
REFERENCE  
AUTHORS 11042159  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system-384-format  
JOURNAL sequencing pipeline with 384 multichipillary sequencer  
PUBMED Genome Res. 10 (11), 1757-1771 (2000)  
REFERENCE 11076861 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
REFERENCE  
AUTHORS 5  
The PANTOM Consortium, the RIKEN Genome Exploration Research Group  
Phase I and II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
JOURNAL of 60,770 full-length cDNAs  
PUBMED Nature 420, 563-573 (2002)  
REFERENCE  
AUTHORS 6  
RIKEN Genome Exploration Research Group, Genome Science Group  
(Genome Network Core Team) and the PANTOM Consortium.  
TITLE Antisense Transcription in the Mammalian Transcriptome  
JOURNAL Science 309, 1564-1566 (2005)  
REFERENCE  
AUTHORS 7  
The PANTOM Consortium, Riken Genome Exploration Research Group and  
Genome Science Group (Genome Network Project Core Group).  
TITLE The Transcriptional Landscape of the Mammalian Genome  
JOURNAL Science 309, 1559-1563 (2005)  
REFERENCE  
AUTHORS 8 (bases 1 to 2553)  
Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, T., Kasukawa, T.,  
Kato, H., Kawahara, C., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koyama, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,  
URL:htp://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
COMMENT  
CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:htp://genome.gsc.riken.jp/  
URL:htp://fantom.gsc.riken.jp/  
Location/Qualifiers  
FEATURES  
source  
1.2553  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:D930008124"  
/db\_xref="taxon:10090"  
/db\_xref="taxon:10090"  
/clone="D930008124"  
/issue\_type="head"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="15 days embryo"  
misc\_feature 1.2553  
/note="putative  
unclassifiable"

ORIGIN  
Query Match 6.9%; Score 38.4; DB 6; Length 2553;  
Best Local Similarity 57.5%; Pred. No. 20;  
Matches 69; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 410 CTGACGAAATGTTTATTAACAGAGCTCTAGTATCTTACCGTGTGACCAATCTGA 469  
DB 714 CTGTAGGAATTTCTGTTACAGCTGTTATATCTTTTCAATATGATCTACTATCTGC 773



Qy 470 TCAATACCACTGTGACTGACTAAATTTTCACTGCTTCCGAGTGTCTCTTCACTG 529  
Db 774 AAATAAGATATTGTGCTTCTTCCCTTCCCACTTCTATCCCTGTGTCTCTTCACTG 833.

Search completed: September 16, 2006, 17:03:14  
Job time : 3903 secs

This Page Blank (uspto)

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

# OM protein - protein search, using sw model

Run on: September 16, 2006, 06:14:30 ; Search time 197 Seconds  
(without alignments)  
429.366 Million cell updates/sec

Title: US-10-815-774-6

Perfect score: 991

Sequence: 1 LASVIPDVATINSLEFNQIKN.....FSLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

A\_Geneseq 8:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*
- 10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	991	100.0	185	4 AAB84386	Aab84386 Amino aci
2	991	100.0	320	4 AAB84387	Aab84387 Amino aci
3	991	100.0	320	9 ABE01516	Aee01516 Corynefor
4	874	88.2	185	3 AAY44582	Aay44582 Chryseoba
5	874	88.2	319	3 AAY44583	Aay44583 Chryseoba
6	125.5	12.7	274	9 AEB40953	Aeb40953 L. pneumo
7	125.5	12.7	286	4 AEB37640	Aeb37640 L. pneumo
8	108	10.9	20	4 AAB84385	Aab84385 Internal
9	96	9.7	20	4 AAB84384	Aab84384 N-termina
10	92	9.3	571	7 ABO69131	Aboc69131 Pseudomon
11	91	9.2	594	3 AAG52479	Aag52479 Arabidops
12	91	9.2	594	3 AAG52451	Aag52451 Arabidops
13	91	9.2	594	5 AAB91568	Abb91568 Herbicida
14	90.5	9.1	292	8 ADY09066	Ady09066 Plant ful
15	89.5	9.0	927	8 ADN23916	Adn23916 Bacterial
16	89.5	9.0	927	8 ADN23918	Adn23918 Bacterial
17	86	8.7	1219	7 AAE37900	Aae37900 Rice homo
18	85.5	8.6	442	8 ADX72461	Adx72461 Plant ful
19	85.5	8.6	477	4 AAB30815	Aab30815 Amino aci
20	85.5	8.6	477	10 AEE28467	Aee28467 Yeast unc
21	85.5	8.6	477	10 AEE72832	Aee72832 S. cerevi
22	85.5	8.6	961	8 ADY3617	Ady3617 Plant ful
23	85.5	8.6	1142	8 ADY06625	Ady06625 Plant ful

## ALIGNMENTS

24	85	8.6	264	3 AAG25396	Aag25396 Arabidops
25	85	8.6	292	3 AAG25395	Aag25395 Arabidops
26	85	8.6	294	3 AAG25394	Aag25394 Arabidops
27	85	8.6	594	10 AEF80099	Aef80099 A. thalia
28	85	8.6	907	10 AEE60297	Aee60297 Cat chlam
29	85	8.6	907	10 AEE60508	Aee60508 Cat chlam
30	84	8.5	3726	4 ABB63947	Abb63947 Drosophila
31	83.5	8.4	459	9 ABB93408	Abb93408 M. xanthu
32	82	8.3	703	8 ADA44005	Ada44005 Bacterial
33	82	8.3	927	8 AAY35054	Aay35054 Chlamydia
34	82	8.3	928	2 AAM88423	Aam88423 Chlamydia
35	82	8.3	928	3 AAY90239	Aay90239 Chlamydia
36	82	8.3	928	5 ABB90542	Abb90542 Chlamydia
37	82	8.3	928	9 ADM65028	Adm65028 C. pneumo
38	82	8.3	929	9 AEC95710	Aec95710 C. pneumo
39	81.5	8.2	2360	6 ABB56972	Abb56972 E. maxima
40	81	8.2	270	3 AAG21040	Aag21040 Arabidops
41	81	8.2	304	3 AAG21039	Aag21039 Arabidops
42	81	8.2	346	4 ABB71526	Abb71526 Drosophila
43	81	8.2	486	3 AAG21038	Aag21038 Arabidops
44	81	8.2	623	8 ADP04503	Adp04503 Sea squir
45	81	8.2	1137	2 AAR42081	Aar42081 Impatiens

RESULT 1  
ID AAB84386 standard; protein; 185 AA.  
XX  
AC AAB84386;  
XX  
DT 06-AUG-2003 (revised)  
DT 22-AUG-2001 (first entry)  
XX  
DE Amino acid sequence of a protein-deamidating enzyme.  
XX  
KW Protein-deamidating enzyme; mineral absorption; food allergy; dough;  
KW bakery; confectionery.  
XX  
OS Chryseobacterium sp.  
XX  
PN EP1106696-A1.  
XX  
PD 13-JUN-2001.  
XX  
PF 04-DEC-2000; 2000EP-00310768.  
XX  
PR 03-DEC-1999; 99JP-00345044.  
XX  
PA (AMANO-) AMANO ENZYME INC.  
XX  
PI Yamaguchi S;  
XX  
DR WPI: 2001-376907/40.  
DR N-PSDB; AAF90280.  
XX  
PT New enzyme for use in e.g. bakery has an ability to deamidate amido  
PT groups in a protein.  
XX  
PS Claim 4; Page 22; 43pp; English.  
XX  
CC The present sequence represents a protein-deamidating enzyme from  
CC Cryeobacterium sp. number 9670. The enzyme is able to deamidate amido  
CC groups in a protein by directly acting upon the amido groups without  
CC cutting peptide bonds and without cross-linking the protein. The enzyme  
CC thus reduces the mineral sensitivity of the protein and increases the  
CC soluble mineral content in the protein-mineral solution, improving the  
CC absorption of minerals in the human body. The enzyme reduces or removes  
CC the toxicity and allergenic property of the protein in a food e.g.  
CC allergy. The enzyme is useful for the improvement of dough in the field  
CC of bakery and confectionery e.g. for the production of crackers,

CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean  
 CC protein products, in various food articles e.g. meat or fish products and  
 CC noodles; and for improving functionality of plant or animal protein.  
 CC (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 185 AA;

Query Match 100.0%; Score 991; DB 4; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-94;  
 Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LASVTPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDGCVARAHKROILMNNGYDC 60  
 DB 1 LASVTPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDGCVARAHKROILMNNGYDC 60  
 QY 61 EKQFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSSGPVDTAMRNA 120  
 DB 61 EKQFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSSGPVDTAMRNA 120  
 QY 121 CVNTSCGSASVSYANTAGNVYRSPNSYLYDNNLINTNCVLTFRSLSGCSPSPAPDV 180  
 DB 121 CVNTSCGSASVSYANTAGNVYRSPNSYLYDNNLINTNCVLTFRSLSGCSPSPAPDV 180  
 QY 181 SSCGF 185  
 DB 181 SSCGF 185

# RESULT 2

AAB84387  
 ID AAB84387 standard; protein; 320 AA.

XX AAB84387;

DT 06-AUG-2003 (revised)  
 DT 22-AUG-2001 (first entry)

DE Amino acid sequence of a protein-deamidating enzyme.

KM Protein-deamidating enzyme; mineral absorption; food allergy; dough;  
 KM bakery; confectionery.

OS Chryseobacterium sp.

PH Key Location/Qualifiers  
 FT Region 1..135  
 XX /note= "prepro region"

PN EPI106696-A1.

PD 13-JUN-2001.

PF 04-DEC-2000; 2000EP-00310768.

PR 03-DEC-1999; 99JP-00345044.

PA (AMANO-) AMANO ENZYME INC.

PI Yamaguchi S;

DR WPI; 2001-376907/40.

DR N-PSDB; MAF90281.

PT New enzyme for use in e.g. bakery has an ability to deamidate amido  
 PT groups in a protein.

PS Example 11; Page 24; 43pp; English.

XX The present sequence represents a protein-deamidating enzyme from  
 CC Cryseobacterium sp. number 9670. The enzyme is able to deamidate amido  
 CC groups in a protein by directly acting upon the amido groups without  
 CC cutting peptide bonds and without cross-linking the protein. The enzyme  
 CC thus reduces the mineral sensitivity of the protein and increases the

CC soluble mineral content in the protein-mineral solution, improving the  
 CC absorption of minerals in the human body. The enzyme reduces or removes  
 CC the toxicity and allergenic property of the protein in a food e.g.  
 CC allergy. The enzyme is useful for the improvement of dough in the field  
 CC of bakery and confectionery e.g. for the production of crackers,  
 CC biscuits, cookies, pizza pies or crusts of pie; in producing soybean  
 CC protein products, in various food articles e.g. meat or fish products and  
 CC noodles; and for improving functionality of plant or animal protein.  
 CC (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 320 AA;

Query Match 100.0%; Score 991; DB 4; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-93;  
 Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LASVTPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDGCVARAHKROILMNNGYDC 60  
 DB 136 LASVTPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDGCVARAHKROILMNNGYDC 195  
 QY 61 EKQFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSSGPVDTAMRNA 120  
 DB 196 EKQFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSSGPVDTAMRNA 255  
 QY 121 CVNTSCGSASVSYANTAGNVYRSPNSYLYDNNLINTNCVLTFRSLSGCSPSPAPDV 180  
 DB 256 CVNTSCGSASVSYANTAGNVYRSPNSYLYDNNLINTNCVLTFRSLSGCSPSPAPDV 315  
 QY 181 SSCGF 185  
 DB 316 SSCGF 320

# RESULT 3

AEE01516  
 ID AEE01516 standard; protein; 320 AA.

XX AEE01516;

DT 26-JUN-2006 (first entry)

DE Coryneform protein production system-related protein, SEQ ID NO:4.

KM protein production; cell culture; genetically engineered microorganism.

OS Unidentified.

PN WO2005103278-A1.

PD 03-NOV-2005.

PF 20-APR-2005; 2005WO-JP007518.

PR 20-APR-2004; 2004JP-00124196.

PR 13-JAN-2005; 2005JP-00005896.

PA (AJIN ) AJINOMOTO CO INC.

PI Date M, Kikuchi Y, Itaya H, Nakamura N;

DR WPI; 2005-748042/76.

DR N-PSDB; AEE01515.

PT Producing heterologous protein by culturing Coryneform bacterium having  
 PT expression gene construct comprising promoter sequence, nucleic acid  
 PT encoding Tat system-dependent signal peptide region and nucleic acid  
 PT encoding foreign protein.

PS Disclosure; SEQ ID NO 4; 47pp; Japanese.

XX The new invention relates to a method of producing heterologous protein,  
 CC by culturing a Coryneform bacterium having an expression gene construct  
 CC carrying, in the direction from the 5'-terminal to 3'-terminal, a

CC promoter sequence functioning in the bacterium, a nucleic acid sequence  
 CC encoding a Tat system-dependent signal peptide region and a nucleic acid  
 CC sequence encoding a foreign protein, and allowing the bacterium to  
 CC produce and secrete the protein. The signal peptide of (M1) comprises SEQ  
 CC ID NO: 28-33. The signal peptide is a signal peptide of isomaltolactonase,  
 CC which comprises SEQ ID NO: 6, or of trimethylamine N-oxidoreductase,  
 CC which comprises SEQ ID NO: 8. The bacteria comprises gene that encodes  
 CC Tat system secretion peptide such as tatA, tatB, tatC or tatE. (M1) is  
 CC useful for producing heterologous protein e.g. glucanase or  
 CC isomaltodextran. (M1) enables efficient production of heterologous  
 CC protein. Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from Wipo at ftp.wipo.int/pub/published.pat.sequences. The present  
 CC sequence is a protein related to the invention.

XX SQ Sequence 320 AA;

Query Match 100.0%; Score 991; DB 9; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-93;  
 Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LASVDPVATLNSLFNQIKNOSCGTSTASSPCITFRYPVDCYARAHKMQILMANGYDC 60  
 DB 136 LASVDPVATLNSLFNQIKNOSCGTSTASSPCITFRYPVDCYARAHKMQILMANGYDC 195  
 QY 61 EKQFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSGGPVTDTAMRNA 120  
 DB 196 EKQFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSGGPVTDTAMRNA 255  
 QY 121 CVNTSCGASVSSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFLSGCSPSPAPDV 180  
 DB 256 CVNTSCGASVSSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFLSGCSPSPAPDV 315  
 QY 181 SSCGF 185  
 DB 316 SSCGF 320

#### RESULT 4

AAV44582  
 ID AAV44582 standard; protein; 185 AA.

AC AAV44582;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 04-APR-2000 (first entry)

DE Chryseobacterium gleum protein-deamidating enzyme.

XX Protein deamidating enzyme; soil bacterium; deamidation activity;  
 KW amido group; carboxyl group; ammonia; cross linking; peptide bond;  
 KW protein engineering; surface hydrophobicity; toxicity; allergic;  
 KW mineral sensitivity; calcium; absorption; mineral enhancing agent.

XX Chryseobacterium gleum; 'JCM 2410'.

OS EP976829-A2.

XX EP976829-A2.

XX 02-FEB-2000.

XX 04-JUN-1999; 99EP-00304367.

XX 04-JUN-1998; 98JP-00173940.

XX (AMAN) AMANO PHARM KK.

XX Yamaguchi S, Matsuura A;

XX WPI: 2000-118552/11.

XX N-PSDB; AA249494.

PT New enzyme for modifying and improving the function of proteins and/or  
 PT peptides has deamidating activity without causing cross linking.

XX PS Claim 5; Page 24; 57pp; English.

XX The present amino acid sequence is the protein-deamidating enzyme,  
 CC isolated from a new strain of soil bacterium, Chryseobacterium gleum JCM  
 CC 2410. This enzyme exerts the deamidation activity by directly acting upon  
 CC side chain amido groups in the protein in bonded state and releasing side  
 CC chain carboxyl groups and ammonia. It can deaminate high molecular weight  
 CC proteins, without cross linking and cleavage of peptide bonds, to improve  
 CC protein function. This sequence is used for protein engineering, to cause  
 CC an increase in surface hydrophobicity and improve the function of a plant  
 CC or animal protein. It can also be used to remove or reduce toxicity or  
 CC allergic property of proteins in food, decrease mineral sensitivity of  
 CC protein, to allow greater absorption into the body and to solubilise  
 CC calcium for use in drinks and mineral enhancing agents. (Updated on 12-  
 CC SEP-2003 to standardise OS field)

XX SQ Sequence 185 AA;

Query Match 88.2%; Score 874; DB 3; Length 185;  
 Best Local Similarity 87.4%; Pred. No. 8.3e-82;  
 Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 3 SVTPVATLNSLFNQIKNOSCGTSTASSPCITFRYPVDCYARAHKMQILMANGYDCXK 62  
 DB 3 SVTPVATLNSLFNQIKNOSCGTSTASSPCITFRYPVDCYARAHKMQILMANGYDCXK 62  
 QY 63 QFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSGGPVTDTAMRNA 122  
 DB 63 QFYVGNLKASTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSGGPVTDTAMRNA 122  
 QY 123 NTSCGASVSSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFLSGCSPSPAPDVSS 182  
 DB 123 NTSCGASVSSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFLSGCSPSPAPDVSS 182  
 QY 183 CGF 185  
 DB 183 CGF 185

#### RESULT 5

AAV44583  
 ID AAV44583 standard; protein; 319 AA.

AC AAV44583;  
 XX

DT 12-SEP-2003 (revised)  
 DT 04-APR-2000 (first entry)

DE Chryseobacterium gleum protein-deamidating prepro-enzyme.

XX Protein deamidating prepro-enzyme; soil bacterium; deamidation activity;  
 KW amido group; carboxyl group; ammonia; cross linking; peptide bond;  
 KW protein engineering; surface hydrophobicity; toxicity; allergic;  
 KW mineral sensitivity; calcium; absorption; mineral enhancing agent.

XX Chryseobacterium gleum; 'JCM 2410'.

OS EP976829-A2.

XX EP976829-A2.

XX 02-FEB-2000.

XX 04-JUN-1999; 99EP-00304367.

XX 04-JUN-1998; 98JP-00173940.

XX (AMAN) AMANO PHARM KK.

XX Yamaguchi S, Matsuura A;

XX WPI: 2000-118552/11.

XX N-PSDB; AA249494.

PT New enzyme for modifying and improving the function of proteins and/or  
 PT peptides has deamidating activity without causing cross linking.

PD 02-FEB-2000.  
 XX 04-JUN-1999; 99EP-00304367.  
 PF 04-JUN-1998; 98JP-00173940.  
 XX (AMANO ) AMANO PHARM KK.  
 XX Yamaguchi S, Matsuura A;  
 PI WPI; 2000-118552/11.  
 DR N-PSDB; AAZ49495.  
 XX  
 PT New enzyme for modifying and improving the function of proteins and/or  
 PT peptides has deamidating activity without causing cross linking.  
 PS Example 26; Page 26; 57pp; English.  
 XX  
 CC The present amino acid sequence is the protein-deamidating prepro-enzyme,  
 CC isolated from a new strain of soil bacterium, *Chryseobacterium gleum* JCM  
 CC 2410. This enzyme exerts the deamidation activity by directly acting upon  
 CC side chain amido groups in the protein in bonded state and releasing side  
 CC chain carboxyl groups and ammonia. It can deaminate high molecular weight  
 CC proteins, without cross linking and cleavage of peptide bonds, to improve  
 CC protein function. This sequence is used for protein engineering, to cause  
 CC an increase in surface hydrophobicity and improve the function of a plant  
 CC or animal protein. It can also be used to remove or reduce toxicity of  
 CC allergenic property of proteins in food, decrease mineral sensitivity of  
 CC protein, to allow greater absorption into the body and to solubilise  
 CC calcium for use in drinks and mineral enhancing agents. (Updated on 12-  
 CC SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 319 AA;

Query Match 88.2%; Score 874; DB 3; Length 319;  
 Best Local Similarity 87.4%; Pred. No. 1.7e-81;  
 Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

```

QY 3 SVIPDVATINSLEFNOIKNOSCGTSTASPCITFRYPVDCGYARAHKRRQILMANGYDCEK 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 137 SVIPDLATINSLFQIKNQAQGTSTASPCITFRYPVDCGYARAHKRRQILMANGYDCEK 196
QY 63 QFYVGNLKAATGTCVAMSHYVALIVSYKASGVTERRIIDPSLFSSGPTDPTAMRWACV 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 197 QFYVGNLRASVGTCCVSWYHVALIVSKNAGIVERRIIDPSLFSSGPTDPTAMRWACT 236
QY 123 NTSGGSASVSSYANTAGNVRYSNSLYDNMLINTNCVLTKEFSLISGSPSPAPDVSS 182
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 257 NTSGGSASVSSYANTAGNVRYSNSLYDNMLINTNCVLTKEFSLISGSPSPAPDVSS 316
QY 183 CGF 185
   |||||
DB 317 CGF 319
```

RESULT 6  
 AEB40953  
 ID AEB40953 standard; protein; 274 AA.  
 XX  
 AC AEB40953;  
 XX  
 DT 08-SEP-2005 (first entry)  
 XX  
 DE *L. pneumophila* protein SEQ ID NO 5285.  
 XX  
 KM detection; infection; Antibacterial; Vaccine.  
 XX  
 OS *Legionella pneumophila*.  
 XX  
 PN WO2005049642-A2.  
 XX  
 PD 02-JUN-2005.  
 XX

PF 23-SEP-2004; 2004WO-IB003578.  
 XX  
 PR 21-NOV-2003; 2003FR-00013687.  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (UTLY-) UNIV LYON 1 BERNARD CLAUDE.  
 PA (CNRS ) CNRS CENT NAT RECH SCT.  
 XX  
 XX Buchrieser C, Tichit M, Etienne J, Ma L, Cazalec C, Glaeser P;  
 PI Rusniok C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;  
 PI Jarraud S;  
 XX  
 DR WPI; 2005-388305/40.  
 XX  
 PT New genome of *Legionella pneumophila* Paris strain and derived  
 PT polypeptides, useful for detection or identification of the strain and  
 PT for treatment and prevention of infections.  
 PS Claim 3; SEQ ID NO 5285; 660pp; English.

XX  
 CC The invention relates to an isolated or purified nucleotide sequences (I)  
 CC from *Legionella pneumophila* Paris strain. (I), and their related  
 CC sequences or fragments, are useful as primers and probes for detection  
 CC and amplification, including differentiation between the Paris and  
 CC Philadelphia strains of *Legionella pneumophila* and to prepare recombinant  
 CC (hybrid) polypeptides (II). (II) are also useful for preparation of  
 CC specific antibodies (Ab), also used for detection/identification of  
 CC *Legionella*, and some (I), specifically those involved in synthesis of  
 CC surface proteins, are targets for identification of inhibitors. (II), or  
 CC vectors that contain (I), are useful as vaccines and immunogenic  
 CC compositions, for treatment and prevention of infections by *L.*  
 CC *pneumophila*. The present sequence represents the amino acid sequence of a  
 CC *L. pneumophila* protein.  
 XX  
 SQ Sequence 274 AA;

Query Match 12.7%; Score 125.5; DB 9; Length 274;  
 Best Local Similarity 31.2%; Pred. No. 0.00035;  
 Matches 43; Conservative 22; Mismatches 36; Indels 37; Gaps 8;

```

QY 3 SVIPDVAT--INSLEFNOIKNOSCGTSTASP-----CITFRYPVDCGYARAHKRRQI 52
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 72 STVPKVSYSDELTMNMLIRD---TRFLYSPDKDPQRRISWLYPDGCFRRLASRLK 127
QY 53 LMMNGYDCE-----KQFYGNLK-----ASTGTCVAMSHYVALIVSYKASGVTERRI 101
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 128 L-----DSHFVYIPAKIFAFGDEMOTPYSSBGT--VSMWYHVASAVNTMGTI-----YV 175
QY 102 IDPSLFSSGPTDPTAMRN 119
   |||||:|||||
DB 176 LDPRAKPEBGMULDWYN 193
```

RESULT 7  
 AEB37640  
 ID AEB37640 standard; protein; 286 AA.  
 XX  
 AC AEB37640;  
 XX  
 DT 08-SEP-2005 (first entry)  
 XX  
 DE *L. pneumophila* protein SEQ ID NO 1972.  
 XX  
 KM detection; infection; Antibacterial; Vaccine.  
 XX  
 OS *Legionella pneumophila*.  
 XX  
 PN WO2005049642-A2.  
 XX  
 PD 02-JUN-2005.  
 XX  
 PF 23-SEP-2004; 2004WO-IB003578.

```
XX 21-NOV-2003; 2003FR-00013687.
XX
XX (INSP ) INST PASTEUR.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (UNLY-) UNIV LYON I BERNARD CLAUDE.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Buchrieser C, Ticht M, Etienne J, Ma L, Cazalot C, Glaeser P,
XX Ruenick C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F,
XX Jarraud S;
XX
XX WPI; 2005-388305/40.
XX
XX New genome of Legionella pneumophila Paris strain and derived
XX polypeptides, useful for detection or identification of the strain and
XX for treatment and prevention of infections.
XX
XX Claim 3; SEQ ID NO 1972; 660pp; English.
XX
XX The invention relates to an isolated or purified nucleotide sequences (I)
XX from Legionella pneumophila Paris strain (I), and their related
XX sequences or fragments, are useful as primers and probes for detection
XX and amplification, including differentiation between the Paris and
XX Philadelphia strains of Legionella pneumophila and to prepare recombinant
XX (hybrid) polypeptides (II). (II) are also useful for preparation of
XX specific antibodies (Ab), also used for detection/identification of
XX Legionella, and some (I), specifically those involved in synthesis of
XX surface proteins, are targets for identification of inhibitors. (II), or
XX vectors that contain (I), are useful as vaccines and immunogenic
XX compositions, for treatment and prevention of infections by L.
XX pneumophila. The present sequence represents the amino acid sequence of a
XX L. pneumophila protein.
XX
XX Sequence 286 AA;
XX
XX Query Match 12.7%; Score 125.5; DB 9; Length 286;
XX Best Local Similarity 31.2%; Pred. No. 0.00037;
XX Matches 43; Conservative 22; Mismatches 36; Indels 37; Gaps 8;
XX
XX 3 SVIPVAT--LNSLPQIKNOSCGTSTAASP-----CTFRYPVDCVRAHMKROI 52
XX | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
XX 84 STVPVSSYDELTMNMLIRD---TRFLYSPDKDPFORIRISMLYPDDGCFARALSRIX 139
XX
XX 53 LMNNGVDC-----KQFYGNLK-----ASTGTCVMSYVATILNSYKXNSGVTEKRI 101
XX 140 L-----DSHFVYIPAKIFAFGLDEMOTPYSSGCT--VSMWYHVASVNNYMTI-----YV 187
XX
XX 102 IDPSLFFSGPVTDTFAMRN 119
XX : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
XX 168 LDPAAKPEGPMIDIMWYN 205
XX
XX
XX RESULT 8
XX AAB84385
XX ID AAB84385 standard; peptide; 20 AA.
XX
XX AAB84385;
XX
XX 06-AUG-2003 (revised)
XX DT 22-AUG-2001 (first entry)
XX
XX Internal peptide of protein-deamidating enzyme.
XX
XX Protein-deamidating enzyme; mineral absorption; food allergy; dough;
XX bakery; confectionery.
XX
XX Chryseobacterium sp.
XX
XX EPI106696-A1.
XX
XX 13-JUN-2001.
XX
XX
```

```
PF 04-DEC-2000; 2000EP-00310768.
XX
XX 03-DEC-1999; 99JP-00345044.
XX
XX (AMAN-) AMANO ENZYME INC.
XX
XX Yamaguchi S;
XX
XX WPI; 2001-376907/40.
XX
XX New enzyme for use in e.g. bakery has an ability to deamidate amido
XX groups in a protein.
XX
XX Example 11; Page 20; 43pp; English.
XX
XX The present sequence represents a peptide fragment of a protein-
XX deamidating enzyme from Chryseobacterium sp. number 9670. The enzyme is
XX able to deamidate amido groups in a protein by directly acting upon the
XX amido groups without cutting peptide bonds and without cross-linking the
XX protein. The enzyme thus reduces the mineral sensitivity of the protein
XX and increases the soluble mineral content in the protein-mineral
XX solution, improving the absorption of minerals in the human body. The
XX enzyme reduces or removes the toxicity and allergenic property of the
XX protein in a food e.g. allergy. The enzyme is useful for the improvement
XX of dough in the field of bakery and confectionery e.g. for the production
XX of crackers, biscuits, cookies, pizza pies or crusts of pie; in producing
XX soybean protein products, in various food articles e.g. meat or fish
XX products and noodles; and for improving functionality of plant or animal
XX protein. (Updated on 06-AUG-2003 to correct OS field.)
XX
XX Sequence 20 AA;
XX
XX Query Match 10.9%; Score 108; DB 4; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 0.00066;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 145 SPNSNYLYDNMLINTNCVLT 164
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1 SPNSNYLYDNMLINTNCVLT 20
XX
XX
XX RESULT 9
XX AAB84384
XX ID AAB84384 standard; peptide; 20 AA.
XX
XX AAB84384;
XX
XX 06-AUG-2003 (revised)
XX DT 22-AUG-2001 (first entry)
XX
XX N-terminal peptide of protein-deamidating enzyme.
XX
XX Protein-deamidating enzyme; mineral absorption; food allergy; dough;
XX bakery; confectionery.
XX
XX Chryseobacterium sp.
XX
XX EPI106696-A1.
XX
XX 13-JUN-2001.
XX
XX 04-DEC-2000; 2000EP-00310768.
XX PF 03-DEC-1999; 99JP-00345044.
XX
XX (AMAN-) AMANO ENZYME INC.
XX
XX Yamaguchi S;
XX
XX WPI; 2001-376907/40.
XX
XX New enzyme for use in e.g. bakery has an ability to deamidate amido
XX groups in a protein.
XX
XX
```

XX Example 11; Page 20; 43pp; English.

XX The present sequence represents a peptide fragment of a protein-

CC deamidating enzyme from *Cryseobacterium* sp. number 9670. The enzyme is

CC able to deamidate amide groups in a protein by directly acting upon the

CC amide groups without cutting peptide bonds and without cross-linking the

CC protein. The enzyme thus reduces the mineral sensitivity of the protein

CC and increases the soluble mineral content in the protein-mineral

CC solution, improving the absorption of minerals in the human body. The

CC enzyme reduces or removes the toxicity and allergenic property of the

CC protein in a food e.g. allergy. The enzyme is useful for the improvement

CC of dough in the field of bakery and confectionery e.g. for the production

CC of crackers, biscuits, cookies, pizza pies or crusts of pie; in producing

CC soybean protein products, in various food articles e.g. meat or fish

CC products and noodles; and for improving functionality of plant or animal

CC protein. (Updated on 06-AUG-2003 to correct OS field.)

XX

SO Sequence 20 AA;

Query Match 9.7%; Score 96; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.012;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LASVIPVATLNSLLENOIKN 20

1 LASVIPVATLNSLLENOIKN 20

Db 1 LASVIPVATLNSLLENOIKN 20

RESULT 10

ID ABO69131 standard; protein; 571 AA.

XX ABO69131;

AC ABO69131;

XX 29-JUL-2004 (first entry)

DT 29-JUL-2004 (first entry)

XX

DE Pseudomonas aeruginosa polypeptide #1306.

XX

KM Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX

OS Pseudomonas aeruginosa.

XX

PN US6551795-B1.

XX

PD 22-APR-2003.

XX

PF 18-FEB-1999; 99US-00252991.

XX

PR 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX

DR WPI; 2003-615309/58.

DR N-PSDB; ABD02702.

XX

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,

PT useful as molecular targets for diagnostics, prophylaxis and treatment of

PT pathological conditions resulting from bacterial infection.

XX

PS Disclosure; SEQ ID NO 17877; 455bp; English.

XX

CC The invention relates to Pseudomonas aeruginosa polypeptides and the

CC polynucleotides encoding them. The sequences are useful in diagnosis and

CC therapy of pathological conditions, as molecular targets for diagnostics,

CC prophylaxis and treatment of pathological conditions resulting from a

CC bacterial infection, for evaluating a compound, such as a polypeptide,

CC for the ability to bind a P. aeruginosa nucleic acid, as components of

CC effective antibacterial targets, as targets for antibacterial drugs,

CC including anti-P. aeruginosa drugs, as templates for recombinant

CC production of P. aeruginosa-derived peptides or polypeptides, as target

CC components for diagnosis and/or treatment of P. aeruginosa-caused

CC infection, and in detection of P. aeruginosa sequences or other sequences

CC of Pseudomonas species using biochip technology. Sequences ABO67826-

CC ABO44396 represent P. aeruginosa polypeptides of the invention. Note: The

CC sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format from USPTO at

CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)

XX

SO Sequence 571 AA;

Query Match 9.3%; Score 92; DB 7; Length 571;

Best Local Similarity 21.6%; Pred. No. 2.7;

Matches 42; Conservative 24; Mismatches 54; Indels 74; Gaps 9;

QY 15 ENOINNOGSGSTA--SSPITFRYPVDC--YARAHK-----MROILMNGYDC 60

15 ENOINNOGSGSTA--SSPITFRYPVDC--YARAHK-----MROILMNGYDC 60

Db 294 FTQVR--SCGISTAVSPCMESQYPREDYSDKAKTHGSLDLIDRAGVQVIMLENSDC 351

294 FTQVR--SCGISTAVSPCMESQYPREDYSDKAKTHGSLDLIDRAGVQVIMLENSDC 351

QY 61 EKQFYGNLKAATGTC-----CWMSYHVAIIIVSYRNAGSVTER 99

61 EKQFYGNLKAATGTC-----CWMSYHVAIIIVSYRNAGSVTER 99

Db 352 K-----GTCIRVPRNDIPKIQSPFCGDKNCIDESLVLGL--GEYIDGLQPD 396

352 K-----GTCIRVPRNDIPKIQSPFCGDKNCIDESLVLGL--GEYIDGLQPD 396

QY 100 RII-----DPSLFSSGPEVTDTAMRNACVNTSCGSASVSSYANTAGVYTRFSNSTL 151

100 RII-----DPSLFSSGPEVTDTAMRNACVNTSCGSASVSSYANTAGVYTRFSNSTL 151

Db 397 AIIIVLHSDGSHGPEYERYEKEMERFQPVCRITNOLGSCSKELVN-----V 442

397 AIIIVLHSDGSHGPEYERYEKEMERFQPVCRITNOLGSCSKELVN-----V 442

QY 152 YDNNLINTNCVLT/K 165

152 YDNNLINTNCVLT/K 165

Db 443 YDNTTILYTDHFLTK 456

443 YDNTTILYTDHFLTK 456

RESULT 11

ID AAG52479 standard; protein; 594 AA.

XX AAG52479;

AC AAG52479;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 66714.

XX

KM Protein identification; signal transduction pathway; metabolic pathway;

KM hybridisation assay; genetic mapping; gene expression control; promoter;

KM termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.



[illegible]

```
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161599P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 9.2%; Score 91; DB 3; Length 594;
Best Local Similarity 24.6%; Pred. No. 3.7;
Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;

QY 25 TSTASSPCTIFRY---PVD--GCYARAHKMRQILMNGY---DCEKQFYGNLAKST-- 73
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 442 TIVDDNPSTVFNFYFKHPVDLQRCVEAIRLVSKVTYSNRFLNLYQCDKQNVHKMLSLSYVA 501
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 74 -----GTCGVAMSTYVALVNS-----YKNAAGVTEKRIIDSLPS 108
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 502 NINIRPKQINDTYKSMQFCCKDITVVTIMHYHGCLVKGKVSNNKVLGVDRLRVIDSTPD 561
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 109 GSPVTD 114
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 562 ESEPGTN 567

RESULT 12
ID AAG52451
AAG52451 standard; protein: 594 AA.
AC AAG52451;
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 66675.
DE
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
FN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-00301439.
PF
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
```

```
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 14-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 19-MAY-1999; 99US-0135124P.
PR 20-MAY-1999; 99US-0135353P.
PR 21-MAY-1999; 99US-0135629P.
PR 24-MAY-1999; 99US-0136021P.
PR 25-MAY-1999; 99US-0136392P.
PR 27-MAY-1999; 99US-0136782P.
PR 28-MAY-1999; 99US-0137222P.
PR 01-JUN-1999; 99US-0137528P.
PR 03-JUN-1999; 99US-0137502P.
PR 04-JUN-1999; 99US-0137724P.
PR 07-JUN-1999; 99US-0138094P.
PR 08-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 26-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 15-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
```

PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149802P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-015138P.  
PR 01-SEP-1999; 99US-015130P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.

PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161923P.  
PR 29-OCT-1999; 99US-0162142P.  
  
Query Match 9.2%; Score 91; DB 3; Length 594;  
Best Local Similarity 24.6%; Pred. No. 3.7;  
Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;  
  
QY 25 TSTASPTCTFRY---PVD--GCYARAHKMQILNNNGY---DDEKOFVGNLAKST-- 73  
Db 442 TIVDDNPSTVFNFYFHPVDLQRCVEAIRLVSKVATSNRPLNTQCCKQVHRLSLVKA 501  
QY 74 -----GTCVWASYHVALVS-----YKNASGVTEKRIIDPSLPS 108  
Db 502 NINLRPKQINDTKSMAQFCQDVIIVITWYHGGCLVGVKVPARKVLGVRLAVIDGSTD 561  
QY 109 GGPVTD 114  
Db 562 EBPQTN 567  
  
RESULT 13  
ABR91568  
ID ABR91568 strand; protein; 594 AA.  
XX  
AC ABR91568;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Herbicidally active polypeptide SEQ ID NO 779.  
XX  
KW Herbicidal; plant; agriculture; herbicide.  
OS Arabidopsis thaliana.  
PN WO200210210-A2.  
XX  
PD 07-FEB-2002.  
XX  
PF 28-AUG-2001; 2001WO-EP009892.  
XX  
PR 28-AUG-2001; 2001WO-EP009892.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Tietjen K, Weidier M;  
XX  
DR WPI; 2002-269010/31.  
XX  
PT Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms.  
XX  
PS Claim 5; SEQ ID NO 779; 261bp + Sequence Listing; English.  
XX  
CC The invention relates to identifying target proteins (ABR90790-ABR94016)  
CC for herbicidally active compounds, comprising aligning and comparing  
CC nucleic acid or amino acid sequences from plant with nucleic acid or  
CC amino acid sequences from non-plant organisms using suitable search

CC parameters, where plant sequences having an E-value greater by a factor  
 CC of 3 than the E-value of most similar non-plant sequences are selected.  
 CC The polypeptides or nucleic acids encoding them are useful for  
 CC identifying modulators. The identified modulators are useful as  
 CC herbicides

XX Sequence 594 AA;

Query Match 9.2%; Score 91; DB 5; Length 594;

Best Local Similarity 24.6%; Pred. No. 3.7; Indels 36; Gaps 5;  
 Matches 31; Conservative 14; Mismatches 45;

QY 25 TSNASSPCITFRY--PVD--GCYARAHKROILMNNGY----DCEKQFYGNLKAFT-- 73

DB 442 TNVDDNPSVTYFKRPVLDQRVEALRLSVKVTNSRFLNTQCCKQVHMLSLSYVA 501

QY 74 -----GTCCVAMSYHVALIVS-----YKMSGVTREKRIIDPSLFS 108

DB 502 NINLRPKQLNDTKSMAQFCQDTVTVTIMHYHGCLVGRKVVSPNRKVLGVDRRLVTDGSTPD 561

QY 109 SGPTD 114

DB 562 ESPGTN 567

RESULT 14

ADY09066 standard; protein; 292 AA.

XX ADY09066;

DT 21-APR-2005 (first entry)

XX Plant full length insert polypeptide seqid 64881.

XX plant protectant; plant growth regulant; gene therapy; plant;

KW recombinant DNA construct; physical array; plant breeding marker;

KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;

KW extreme osmotic condition; pathogen tolerance; pest tolerance;

KW growth rate; cell cycle pathway; disease resistance;

KW galactomannan production; lignin production; plant growth regulator;

KW yield; plant growth; plant development; seed oil; protein yield;

XX protein content.

OS Unidentified.

PN US2004034888-A1.

XX 19-FEB-2004.

PF 28-APR-2003; 2003US-00425114.

PR 06-MAY-1999; 99US-00304517.

PA 05-NOV-2001; 2001US-00985678.

PA (LITU/) LITU J.

PA (ZHOU/) ZHOU Y.

PA (KOVA/) KOVALIC D K.

PA (SCRE/) SCREEN S E.

PA (TABAS/) TABASKA J E.

PA (CAOY/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

XX MPI; 2004-180133/17.

XX New recombinant DNA construct, useful for improving plant tolerance to

PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or

PT pests, for conferring increased resistance to plant disease, or for

PT improving yield.

XX Claim 1; SEQ ID NO 64881; 15pp; English.

CC The invention describes a recombinant DNA construct comprising a  
 CC polynucleotide consisting of a sequence encoding an amino acid sequence  
 CC available in electronic form from the US patent office at  
 CC ftp.segdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide  
 CC of the invention are also useful in physical arrays of molecules and as  
 CC plant breeding markers. The recombinant DNA construct is useful for  
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, for conferring  
 CC increased resistance to plant disease, for increasing the rate of homologous  
 CC lignin or plant growth regulators, for improving yield by modification of  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This is the amino acid sequence of a plant full length insert  
 CC polypeptide that can be used in the recombinant DNA construct of the  
 CC invention.

XX Sequence 292 AA;

Query Match 9.1%; Score 90.5; DB 8; Length 292;

Best Local Similarity 20.8%; Pred. No. 1.6; Indels 68; Gaps 10;

Matches 47; Conservative 32; Mismatches 68;

QY 26 STASSPCITFRYPYDG-----CYARAHKROILMNNGYCCEKQFY--YGNLKA 71

DB 31 AVAVAAACAFSPSPYSGIRKIGFIQPIVCRSTVOGRHLISDNGYVSALINPWSHCCP 90

QY 72 STG-----TCCVAMSYHVALIVSVKMSGVTREKRIIDPSLFSGPTDTAMRN 119

DB 91 TTGDRFSCGGCELDLQCCNSYEVCVSCCL--NPSKIKKEDVL--KLKAKKPTAGTYTN 145

QY 120 A-----CVNTSCGSAVSSTSYA-----NTAGNVYSPNSNYLYDNKLI----- 157

DB 146 VDFPCMGKCRHSASVYHENAASDFHYCFVSQNSSGST--EBSVSYSKILGINVVGRP 203

QY 158 -----NTNCVLTKEFSL-----SGCSPSPAPD 179

DB 204 RSCSLVCKVRQSCVPRLSYLNKCEILOKYMCKSCGSPSLGPD 249

RESULT 15

ADN23916 standard; protein; 927 AA.

XX ADN23916;

DT 02-DEC-2004 (first entry)

XX Bacterial polypeptide #6569.

KW Recombinant DNA construct; transformed plant; improved plant property;

KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

KW pathogen tolerance; pest tolerance; plant disease resistance;

KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;

KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

XX bacterial polypeptide.

XX Bacteria.

XX US200323675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.



This Page Blank (uspto)

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2006, 06:23:35 ; Search time 50 Seconds  
(without alignments)  
323.865 Million cell updates/sec

Title: US-10-815-774-6  
Perfect score: 991  
Sequence: 1 LASV1PDVATLNSLFNQIKN.....FSLSGCSPSPAPDVSSCGF 185

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /EMC Celerra\_SIDS3/ptodata/2/1aa/5\_COMB.pep:\*  
2: /EMC Celerra\_SIDS3/ptodata/2/1aa/6\_COMB.pep:\*  
3: /EMC Celerra\_SIDS3/ptodata/2/1aa/7\_COMB.pep:\*  
4: /EMC Celerra\_SIDS3/ptodata/2/1aa/H\_COMB.pep:\*  
5: /EMC Celerra\_SIDS3/ptodata/2/1aa/PTOS\_COMB.pep:\*  
6: /EMC Celerra\_SIDS3/ptodata/2/1aa/RE\_COMB.pep:\*  
7: /EMC Celerra\_SIDS3/ptodata/2/1aa/backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	991	100.0	185	2 US-09-727-769A-6	Sequence 6, Appl
2	991	100.0	320	2 US-09-727-769A-8	Sequence 8, Appl
3	874	88.2	185	2 US-09-324-910-6	Sequence 6, Appl
4	874	88.2	185	2 US-09-793-495-6	Sequence 6, Appl
5	874	88.2	319	2 US-09-324-910-12	Sequence 12, Appl
6	874	88.2	319	2 US-09-793-495-12	Sequence 12, Appl
7	108	10.9	20	2 US-09-727-769A-2	Sequence 1, Appl
8	96	9.7	20	2 US-09-727-769A-1	Sequence 1, Appl
9	92	9.3	571	2 US-09-252-540-17877	Sequence 1, Appl
10	83.5	8.4	459	2 US-09-802-540-12607	Sequence 12607, A
11	83	8.4	20	2 US-09-324-910-2	Sequence 2, Appl
12	83	8.4	20	2 US-09-793-495-2	Sequence 2, Appl
13	82	8.3	927	2 US-09-198-452A-472	Sequence 472, App
14	82	8.3	937	2 US-09-438-185A-449	Sequence 449, App
15	80.5	8.1	710	2 US-09-248-796A-14136	Sequence 14136, A
16	80	8.1	406	2 US-09-248-796A-14345	Sequence 14345, A
17	79	8.0	20	2 US-09-324-910-1	Sequence 1, Appl
18	79	8.0	20	2 US-09-793-495-1	Sequence 1, Appl
19	78	7.9	234	2 US-09-252-991A-29027	Sequence 29027, A
20	78	7.9	357	2 US-09-270-767-45129	Sequence 45129, A
21	77.5	7.8	666	2 US-08-982-785A-11	Sequence 11, Appl
22	77.5	7.8	666	2 US-09-629-498-11	Sequence 11, Appl
23	76.5	7.7	433	2 US-09-104-623A-4	Sequence 4, Appl
24	76.5	7.7	433	2 US-09-019-532-4	Sequence 4, Appl
25	76.5	7.7	433	2 US-09-338-746-4	Sequence 4, Appl
26	76.5	7.7	433	2 US-09-985-689A-3	Sequence 3, Appl

27	76.5	7.7	433	2 US-09-985-689A-4	Sequence 4, Appl
28	76.5	7.7	433	2 US-09-985-689A-5	Sequence 5, Appl
29	76.5	7.7	499	1 US-07-952-853-6	Sequence 6, Appl
30	76.5	7.7	499	1 US-08-914-848-6	Sequence 6, Appl
31	76.5	7.7	635	1 US-08-873-479-43	Sequence 43, Appl
32	76.5	7.7	726	2 US-09-949-016-5916	Sequence 5916, Ap
33	76.5	7.7	737	2 US-09-949-016-11607	Sequence 11607, A
34	76	7.7	1133	2 US-09-902-540-12243	Sequence 12243, A
35	75.5	7.6	149	2 US-09-489-039A-8252	Sequence 8252, Ap
36	75.5	7.6	468	3 US-09-497-967-54	Sequence 54, Appl
37	75.5	7.6	539	2 US-09-902-540-11983	Sequence 11983, A
38	75.5	7.6	1190	2 US-09-949-016-7676	Sequence 7676, Ap
39	75.5	7.6	1190	2 US-09-949-016-7677	Sequence 7677, Ap
40	75.5	7.6	1219	1 US-10-148-806-2	Sequence 2, Appl
41	75	7.6	271	1 US-08-467-155A-10	Sequence 10, Appl
42	75	7.6	271	1 US-08-628-198-10	Sequence 10, Appl
43	75	7.6	271	2 US-09-201-038-10	Sequence 10, Appl
44	75	7.6	271	3 US-10-021-368-10	Sequence 10, Appl
45	75	7.6	271	5 PCT-US96-07343-10	Sequence 10, Appl

## ALIGNMENTS

```
RESULT 1
US-09-727-769A-6
; Sequence 6, Application US/09727769A
; Patent No. 6756221
; GENERAL INFORMATION:
; APPLICANT: Amano Enzyme, Inc.
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, MICROORGANISM PRODUCING THE SA
; FILE OF INVENTION: ENCODING THE SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; CURRENT REFERENCE: 062106
; CURRENT APPLICATION NUMBER: US/09/727,769A
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: JP 11-345044
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; TYPE: PRT
; LENGTH: 185
; ORGANISM: Cryoseobacterium sp. No. 6756221 9670
US-09-727-769A-6

Query Match      100.0%; Score 991; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 5.6e-101;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LASV1PDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNGYDC 60
Db 1 LASV1PDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNGYDC 60
QY 61 EKQFYGNLKAISTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRNA 120
Db 61 EKQFYGNLKAISTGTCCVAMSYHVALIVSYKNASGVTEKRIIDPSLFSSGPTDTAMRNA 120
QY 121 CVNTSCGASVSANTAGNYRRSPNSYLDNMLINTNCTLTCTFSLISGSPSPAPDV 180
Db 121 CVNTSCGASVSANTAGNYRRSPNSYLDNMLINTNCTLTCTFSLISGSPSPAPDV 180
QY 181 SSCGF 185
Db 181 SSCGF 185

RESULT 2
US-09-727-769A-8
; Sequence 8, Application US/09727769A
; Patent No. 6756221
; GENERAL INFORMATION:
; APPLICANT: Amano Enzyme, Inc.
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, MICROORGANISM PRODUCING THE SA
```

```

; TITLE OF INVENTION: ENCODING THE SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 062106
; CURRENT APPLICATION NUMBER: US/09/727,769A
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: JP 11-345044
; PRIOR FILING DATE: 1999-12-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 320
; TYPE: PRF
; ORGANISM: Chryseobacterium sp. No. 6756221 9670
US-09-727-769A-8
```

```

Query Match      100.0%; Score 991; DB 2; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.2e-100;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 LASVDPVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNNGYDC 60
   |||
DB 136 LASVDPVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNNGYDC 195
   |||
QY 61 EKQFYGNLKASTGTCCVAMS YHVALIVSYKNASGVTEKRIIDPSLFSGPTVDTAMRNA 120
   |||
DB 196 EKQFYGNLKASTGTCCVAMS YHVALIVSYKNASGVTEKRIIDPSLFSGPTVDTAMRNA 255
   |||
QY 121 CVNTSCGSASVSSYANTAGNYYRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDV 180
   |||
DB 256 CVNTSCGSASVSSYANTAGNYYRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDV 315
   |||
QY 181 SSCGF 185
   |||
DB 316 SSCGF 320
```

```

RESULT 3
US-09-324-910-6
; Sequence 6, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsura, Akira
; TITLE OF INVENTION: NOVEL, PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 0-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 185
; TYPE: PRF
; ORGANISM: Chryseobacterium gleum
US-09-324-910-6
```

```

Query Match      88.2%; Score 874; DB 2; Length 185;
Best Local Similarity 87.4%; Pred. No. 4.3e-88;
Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;
```

```

QY 3 SVIPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNNGYDC 62
   |||
DB 3 SVIPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNNGYDC 62
   |||
QY 63 QFYGNLKASTGTCCVAMS YHVALIVSYKNASGVTEKRIIDPSLFSGPTVDTAMRNA 122
   |||
DB 63 QFYGNLKASTGTCCVAMS YHVALIVSYKNASGVTEKRIIDPSLFSGPTVDTAMRNA 122
   |||
QY 123 NTS CGSASVSSYANTAGNYYRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDV 182
   |||
DB 123 NTS CGSASVSSYANTAGNYYRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDV 182
   |||
```

```

QY 183 CGF 185
   |||
DB 183 CGF 185
```

```

RESULT 4
US-09-793-495-6
; Sequence 6, Application US/09793495
; Patent No. 6770469
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsura, Akira
; TITLE OF INVENTION: NOVEL, PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 0-54362
; CURRENT APPLICATION NUMBER: US/09/793,495
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: HEI-10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 185
; TYPE: PRF
; ORGANISM: Chryseobacterium gleum
US-09-793-495-6
```

```

Query Match      88.2%; Score 874; DB 2; Length 185;
Best Local Similarity 87.4%; Pred. No. 4.3e-88;
Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;
```

```

QY 3 SVIPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNNGYDC 62
   |||
DB 3 SVIPDVATLNSLFNQIKNOSCGTSTASPCITFRYPVDCGYARAHKROILMNNGYDC 62
   |||
QY 63 QFYGNLKASTGTCCVAMS YHVALIVSYKNASGVTEKRIIDPSLFSGPTVDTAMRNA 122
   |||
DB 63 QFYGNLKASTGTCCVAMS YHVALIVSYKNASGVTEKRIIDPSLFSGPTVDTAMRNA 122
   |||
QY 123 NTS CGSASVSSYANTAGNYYRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDV 182
   |||
DB 123 NTS CGSASVSSYANTAGNYYRSPNSLYLDNNLINTNCVLTFRSLSGCSPAPDV 182
   |||
QY 183 CGF 185
   |||
DB 183 CGF 185
```

```

RESULT 5
US-09-324-910-12
; Sequence 12, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsura, Akira
; TITLE OF INVENTION: NOVEL, PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 0-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 319
; TYPE: PRF
; ORGANISM: Chryseobacterium gleum
US-09-324-910-12
```

```

Query Match      88.2%; Score 874; DB 2; Length 319;
```



	Query March	88.2%	Score 874;	DB 2;	Length 319;	
	Best Local Similarity	87.4%	Pred. No. 9, 3e-88;			
	Matches 160;	Conservative 10;	Mismatches 13;	Indels 0;	Gaps 0	
QY	3 SVIPDVATLNSLFNQIKNKGCGTSTASSPCITFRYPVDGCVARAHKMRQILMNNGYDCBK	62				
Dd	137 SVIPDLATLNLSLFNQIKNACGTSTASSPCITFRYPVDGCVARAHKMRQILNLAGYDCEK	196				
QY	63 QFYVGNLKASNGTCCTCAVSRYHALLIVSYKMGAVTEKRILIDSLFSSGGVTPTARNACY	122				
Dd	197 QFYVGNLKASNGTCCTCAVSRYHALLIVSFKNASGVIEKRILIDSLFSSGGVTDTSAPRAACT	256				
QY	123 NTSCGSASVSSSVANTAGCNVYYRSPENSYLDNNLTINTNCVLTKFSLLSGCSPSPADVSS	182				
Dd	257 NTSCGSASVSSSVANTAGNVYYRSPESGILLYDNNVYNVTNCVLNTFSSLSCSPPSPASVAS	316				
QY	183 CGF 185					
Dd	317 CGF 319					

```

Query Match          9.7%; Score 96; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00088;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 LASVIPDVAATLNSLFNOIKN 20
        |||||
Db       1 LASVIPDVAATLNSLFNOIKN 20

```

```

RESULT 9
US-09-252-991A-17877
; Sequence 17877 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ. ID NO 17877
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-17877

```

```
Query Match          9.3%; Score 92; DB 2; Length 571;
Best Local Similarity 21.6%; Pred. No. 0.29;
Matches 42; Conservative 24; Mismatches 54; Indels 74; Gaps 9;

QY 15 FNOIKNOSGCTGA--SSPITRRYPVDG---YARAHK-----MROILMMNGGYC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 FTQVR--SCGTSTAVSVPCMFQSPYFREDYSDKKAKTHBLDILQRAVQVIMLENSDC 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 EKQFVYGNLKAATGTC-----CVAWSYHAILVSYRNASGVTEK 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 K-----GTLARVPNDIPKTOPSPFCGKNCIDESILVGL---QEVIGLQGD 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 100 RII-----DPSLFSGGPVTDITAMRNACVNTSCGSASVSYANTAGNVYRSPNSYL 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 397 AIIVLHSDSGHGEYERYPKEMERFQPVCRITNQLGSKSEKELVN-----V 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 YDNNLINTNCVLT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 443 YDNTILYTHFLTK 456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-902-540-12607
; Sequence 12607, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12607
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-12607

Query Match          8.4%; Score 83.5; DB 2; Length 459;
Best Local Similarity 23.4%; Pred. No. 1.8;
Matches 47; Conservative 21; Mismatches 60; Indels 73; Gaps 12;

QY 2 ASVTPDVATINSLEN-----QIKNOSGCTSTASSP-CIT---FRYPVDG 41
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 226 ALVVEDVINANALPNTGSGMNEPSGHTFLYQVR---YTTASVPLCPGAMWAVPVKG 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 42 CYARAHKROIIMNNGYDCEKQFYGNLKA-STGTCVAMSYHAILVSYRNASGVTEKRI 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 LFD-----LOGDRYESPRAFTLG-CDAGVIAKCYRWGHPWL-----DCAT----- 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 102 IDSLFSSGPGVTDITAMRNACVNTSCGSASVSYANTAGNVYRSPNSYLDNNLINTNC 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 321 -----SGPTTEAHM-----SCTMRADYD-----GQGTSTFLDGTTRIRPMD 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 162 VITKFSLLSGSPAPDVSS 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 358 ALTP-AIIS---APSPDSS 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-324-910-2
; Sequence 2, Application US/09324910
; Patent No. 6251651
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
```

```
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/324,910
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: HEI-10-173940
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-324-910-2

Query Match          8.4%; Score 83; DB 2; Length 20;
Best Local Similarity 78.9%; Pred. No. 0.024;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 145 SPNSGYLDNNLINTNCVL 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 SPGSLLDNNYVNTNCVL 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-793-495-2
; Sequence 2, Application US/09793495
; Patent No. 6770469
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; APPLICANT: Matsuura, Akira
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
; TITLE OF INVENTION: SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: Q-54362
; CURRENT APPLICATION NUMBER: US/09/793,495
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: HEI-10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-793-495-2

Query Match          8.4%; Score 83; DB 2; Length 20;
Best Local Similarity 78.9%; Pred. No. 0.024;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 145 SPNSGYLDNNLINTNCVL 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 SPGSLLDNNYVNTNCVL 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-198-452A-472
; Sequence 472, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 472
; LENGTH: 927
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
```



**This Page Blank (uspto)**



```
; Sequence 8, Application US/10815774
; Publication No. US20040175799A1
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, MICROORGANISM PRODUCING THE SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 080844
; CURRENT APPLICATION NUMBER: US/10/815,774
; CURRENT FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US/09/727,769
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US/09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: JP Hei. 11-345044
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP Hei. 10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Cryseobacterium sp. No. 9670
US-10-815-774-8
```

Query Match Best Local Similarity 100.0%; Score 991; DB 4; Length 320;

Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 LASVDPVATLNSLFTQIKNOGCGTSTASSPCITFRYPVDCYARAHKMQILMNGYDC 60
    |||||
DB 136 LASVDPVATLNSLFTQIKNOGCGTSTASSPCITFRYPVDCYARAHKMQILMNGYDC 195
    |||||
QY 61 EKQFVGNLKAAGTCCVANSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTAMRNA 120
    |||||
DB 196 EKQFVGNLKAAGTCCVANSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTAMRNA 255
    |||||
QY 121 CVNTSCGSASVSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFSLSGCSPSPADPV 180
    |||||
DB 256 CVNTSCGSASVSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFSLSGCSPSPADPV 315
    |||||
QY 181 SSCGF 185
    |||||
DB 316 SSCGF 320
```

RESULT 3

```
; Sequence 6, Application US/09793495
; Publication No. US20040072318A1
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 0-54362
; CURRENT APPLICATION NUMBER: US/09/793,495
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: HEI-10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-09-793-495-6
```

Query Match Best Local Similarity 87.4%; Score 874; DB 3; Length 185; Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

```
QY 3 SVIPDVATLNSLFTQIKNOGCGTSTASSPCITFRYPVDCYARAHKMQILMNGYDC 62
    |||||
DB 3 SVIPDVATLNSLFTQIKNOGCGTSTASSPCITFRYPVDCYARAHKMQILMNGYDC 62
    |||||
QY 63 QFVGNLKAAGTCCVANSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTAMRNA 122
    |||||
DB 63 QFVGNLKAAGTCCVANSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTAMRNA 122
    |||||
QY 123 NTSOGSASVSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFSLSGCSPSPADPV 182
    |||||
DB 123 NTSOGSASVSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFSLSGCSPSPADPV 182
    |||||
QY 183 CGF 185
    |||||
DB 183 CGF 185
```

RESULT 4

```
US-10-815-751-6
; Sequence 6, Application US/10815751
; Publication No. US2004016658A1
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE SAME, PRODUCTION PROCESS THEREFOR, AND USE THEREOF
; FILE REFERENCE: 0-54362
; CURRENT APPLICATION NUMBER: US/10/815,751
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US/09/793,495
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/324,910
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: HEI-10-173940
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Chryseobacterium gleum
US-10-815-751-6
```

Query Match Best Local Similarity 87.4%; Score 874; DB 4; Length 185;

Matches 160; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

```
QY 3 SVIPDVATLNSLFTQIKNOGCGTSTASSPCITFRYPVDCYARAHKMQILMNGYDC 62
    |||||
DB 3 SVIPDVATLNSLFTQIKNOGCGTSTASSPCITFRYPVDCYARAHKMQILMNGYDC 62
    |||||
QY 63 QFVGNLKAAGTCCVANSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTAMRNA 122
    |||||
DB 63 QFVGNLKAAGTCCVANSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTAMRNA 122
    |||||
QY 123 NTSOGSASVSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFSLSGCSPSPADPV 182
    |||||
DB 123 NTSOGSASVSYANTAGNVYRSPNSLYLDNNLINTNCVLTKEFSLSGCSPSPADPV 182
    |||||
QY 183 CGF 185
    |||||
DB 183 CGF 185
```

RESULT 5

```
US-09-793-495-12
; Sequence 12, Application US/09793495
; Publication No. US20040072318A1
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, Shotaro
; TITLE OF INVENTION: NOVEL PROTEIN-DEAMIDATING ENZYME, GENE ENCODING THE
```













```
APPLICANT: Alexandrov, Nickolai
TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding
FILE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 2750-1590PUS2
CURRENT APPLICATION NUMBER: US/11/056,355B
PRIOR FILING DATE: 2005-02-14
PRIOR APPLICATION NUMBER: 60/544,190
PRIOR FILING DATE: 2004-02-13
NUMBER OF SEQ ID NOS: 119966
SEQ ID NO 83015
LENGTH: 594
TYPE: prt
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: peptide
LOCATION: (1)..(594)
OTHER INFORMATION: Ceres Seq. ID no. 12668218
US-11-056-355B-83015
```

Query Match  
Best Local Similarity 24.6%; Pred. No. 0.87;  
Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;

```
QY 25 TSTASPCITFRY---PVD--GCYARAHKROILMNGY----DCEKQFYGNLKAST-- 73
| : : : : : | : : : : : | : : : : : | : : : : : |
DB 442 TNVDNPSVTENYFKHPVDLQRCVEAIRLVSKVTSNRPLNTQCDKQVHKLSTSVKA 501
QY 74 -----GTCCVMSYHVALIVS-----YKNASGVTEKRIIDPSLFS 108
| : : : : : | : : : : : | : : : : : | : : : : : |
DB 502 NINLRPKQJNDTKSMAQFCQDTVTITWHYHGGLVGVKVSPPNRKVLGVDRLRVIDGSTFD 561
QY 109 SGPVTD 114
| : : : : : |
DB 562 ESPGYN 567
```

RESULT 3  
US-11-056-355B-91487  
; Sequence 91487, Application US/11056355B  
; Publication No. US20060150283A1  
; GENERAL INFORMATION:  
; APPLICANT: Brover, Vyacheslav  
; APPLICANT: Alexandrov, Nickolai  
; TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding  
; FILE OF INVENTION: Polypeptides Encoded Thereby  
; FILE REFERENCE: 2750-1590PUS2  
; CURRENT APPLICATION NUMBER: US/11/056,355B  
; CURRENT FILING DATE: 2005-02-14  
; PRIOR APPLICATION NUMBER: 60/544,190  
; PRIOR FILING DATE: 2004-02-13  
; NUMBER OF SEQ ID NOS: 119966  
; SEQ ID NO 91487  
; LENGTH: 594  
; TYPE: prt  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: (1)..(594)  
; OTHER INFORMATION: Ceres Seq. ID no. 12727965  
US-11-056-355B-91487

Query Match  
Best Local Similarity 24.6%; Pred. No. 0.87;  
Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;

```
QY 25 TSTASPCITFRY---PVD--GCYARAHKROILMNGY----DCEKQFYGNLKAST-- 73
| : : : : : | : : : : : | : : : : : | : : : : : |
DB 442 TNVDNPSVTENYFKHPVDLQRCVEAIRLVSKVTSNRPLNTQCDKQVHKLSTSVKA 501
QY 74 -----GTCCVMSYHVALIVS-----YKNASGVTEKRIIDPSLFS 108
| : : : : : | : : : : : | : : : : : | : : : : : |
DB 502 NINLRPKQJNDTKSMAQFCQDTVTITWHYHGGLVGVKVSPPNRKVLGVDRLRVIDGSTFD 561
```

```
QY 109 SGPVTD 114
| : : : : : |
DB 562 ESPGYN 567
```

RESULT 4  
US-11-056-355B-95243  
; Sequence 95243, Application US/11056355B  
; Publication No. US20060150283A1  
; GENERAL INFORMATION:  
; APPLICANT: Brover, Vyacheslav  
; APPLICANT: Alexandrov, Nickolai  
; TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding  
; FILE OF INVENTION: Polypeptides Encoded Thereby  
; FILE REFERENCE: 2750-1590PUS2  
; CURRENT APPLICATION NUMBER: US/11/056,355B  
; CURRENT FILING DATE: 2005-02-14  
; PRIOR APPLICATION NUMBER: 60/544,190  
; PRIOR FILING DATE: 2004-02-13  
; NUMBER OF SEQ ID NOS: 119966  
; SEQ ID NO 95243  
; LENGTH: 594  
; TYPE: prt  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: (1)..(594)  
; OTHER INFORMATION: Ceres Seq. ID no. 12727965  
US-11-056-355B-95243

Query Match  
Best Local Similarity 24.6%; Pred. No. 0.87;  
Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;

```
QY 25 TSTASPCITFRY---PVD--GCYARAHKROILMNGY----DCEKQFYGNLKAST-- 73
| : : : : : | : : : : : | : : : : : | : : : : : |
DB 442 TNVDNPSVTENYFKHPVDLQRCVEAIRLVSKVTSNRPLNTQCDKQVHKLSTSVKA 501
QY 74 -----GTCCVMSYHVALIVS-----YKNASGVTEKRIIDPSLFS 108
| : : : : : | : : : : : | : : : : : | : : : : : |
DB 502 NINLRPKQJNDTKSMAQFCQDTVTITWHYHGGLVGVKVSPPNRKVLGVDRLRVIDGSTFD 561
QY 109 SGPVTD 114
| : : : : : |
DB 562 ESPGYN 567
```

RESULT 5  
US-10-449-902-42587  
; Sequence 42587, Application US/10449902  
; Publication No. US20060123505A1  
; GENERAL INFORMATION:  
; APPLICANT: National Institute of Agricultural Sciences.  
; APPLICANT: Bio-oriented Technology Research Advancement Institution.  
; APPLICANT: The Institute of Physical and Chemical Research.  
; APPLICANT: Foundation for Advancement of International Science.  
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USBS THEREOF  
; FILE REFERENCE: MOA-A0205YI-US  
; CURRENT APPLICATION NUMBER: US/10/449,902  
; CURRENT FILING DATE: 2003-05-29  
; PRIOR APPLICATION NUMBER: JP 2002-203269  
; PRIOR FILING DATE: 2002-05-30  
; PRIOR APPLICATION NUMBER: JP 2002-383870  
; PRIOR FILING DATE: 2002-12-11  
; NUMBER OF SEQ ID NOS: 36791  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 42587  
; LENGTH: 794  
; TYPE: prt  
; ORGANISM: Oryza sativa  
US-10-449-902-42587

Query Match  
8.7%; Score 86; DB 6; Length 794;



```

; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53325
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-53325

```

```

Query Match      8.5%; Score 84; DB 6; Length 800;
Best Local Similarity 23.6%; Pred. No. 5.7;
Matches 42; Conservative 26; Mismatches 72; Indels 38; Gaps 7;

```

```

QY      12 NSLFGNIGKQSCGTSTASSPCITFRYPVDCYARAHKMQILMNGDCEK-----QPV 65
      215 SSIFRSETTDFNTSMCSP-----SCSRSLSSSIDLHGNGHVEKTSIRVGSFN 265
DB      66 YGNLKAATGT-----CCVAMS-YHVAIIVSYKNAAGVTEKRIIDPSLFGSGPVTDT 115
      266 LRQIKGLSGSMPLPREELRCPISLQMLHDPVII-----ASQGTERRACIEKWFSSGNTTGP 321
QY      116 AMRNACVNTS-----CGSASVSSYANTAGNVYRSPNSG---LYDNNILINTNCVLT 164
DB      322 KTRNELSOLSMPTNYCIRKGLIASWCQNGVLPSPAPDPKLYRIRISSILNISKCLVT 379

```

```

RESULT 10
US-10-530-983-21
; Sequence 21, Application US/10530983
; Publication No. US20060172301A1
; GENERAL INFORMATION:
; APPLICANT: Innovations Foundation
; TITLE OF INVENTION: Secreted Acid Phosphatase (sapw) is Present Only in Pathogenic
; TITLE OF INVENTION: Mycobacteria and Expressed Selectively at Phagosomal pH
; FILE REFERENCE: 4146 0005
; CURRENT APPLICATION NUMBER: US/10/530,983
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: US 60/416,957
; PRIOR FILING DATE: 2002-10-09
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-530-983-21

```

```

Query Match      8.1%; Score 80.5; DB 6; Length 447;
Best Local Similarity 23.1%; Pred. No. 6.2;
Matches 42; Conservative 25; Mismatches 80; Indels 35; Gaps 9;

```

```

QY      33 ITRRYVVDGCVAAHMKROIIMNNGYDCEK-----FVGNLKAATGTCCVA-W 80
DB      244 ILTFPEDGTVSSKSKIFSVLGGALPDELKGTQDDFTYTHYSVIASVSANWGLPSLGRW 303
QY      81 SYHVALIVSYKNSG-----VTEKRIID--PSLFGSGPVT--DTAMRNACVNTSCGSA 129
DB      304 DCGANILEIYANTGYVNVVDFTNRLANFTYGPMSAGSYKXSVWPMALRLRGDCSAG 363
QY      130 S-----VSSYANTAGNVYRSPNSSYLYDN-NLINTNCVLYFSLIS--GCSPPAPDV 180
DB      364 HGILDIIVKETYANTPEYNYSSP---FPYDTASNVTKYATKKNVTGTRHSSSSSSSPSA 420

```

```

QY      181 SS 182
DB      421 SS 422

```

```

RESULT 11
US-11-056-355B-97301
; Sequence 97301, Application US/11056355B
; Publication No. US20060150283A1
; GENERAL INFORMATION:
; APPLICANT: Brover, Vyacheslav
; APPLICANT: Alexandrov, Nikolai
; TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 2750-1590PUS2
; CURRENT APPLICATION NUMBER: US/11/056,355B
; CURRENT FILING DATE: 2005-02-14
; PRIOR APPLICATION NUMBER: 60/544,190
; PRIOR FILING DATE: 2004-02-13
; NUMBER OF SEQ ID NOS: 119966
; SEQ ID NO 97301
; LENGTH: 455
; TYPE: prt
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: (1)..(455)
; OTHER INFORMATION: Ceres Seq. ID no. 12736955
US-11-056-355B-97301

```

```

Query Match      8.1%; Score 80.5; DB 7; Length 455;
Best Local Similarity 29.1%; Pred. No. 6.4;
Matches 37; Conservative 15; Mismatches 46; Indels 29; Gaps 8;

```

```

QY      60 CEKQFVYGNLKAATGTGTC--CVAMS-YHVAIIVSYKNAAGVTEKRIIDPSLFGSGPVTDTAM 117
DB      334 CMLQSLIDPTKYTLGLCHNSQPMSTPONTLTILKXSKLCLE-----STGP----- 378
QY      118 RNACV--NTSCGASVSSYAN-TGANNYY--RSPNSLYLD---NNLINTNCVLTFRS 167
DB      379 -NAPVXLSETSCSSPNLSWETISASNNMLAKSTNNSLCILDVDETNNLMASNCVK-G 436
QY      168 LUSGSCP 174
DB      437 EDSCDP 443

```

```

RESULT 12
US-11-056-355B-97300
; Sequence 97300, Application US/11056355B
; Publication No. US20060150283A1
; GENERAL INFORMATION:
; APPLICANT: Brover, Vyacheslav
; APPLICANT: Alexandrov, Nikolai
; TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 2750-1590PUS2
; CURRENT APPLICATION NUMBER: US/11/056,355B
; CURRENT FILING DATE: 2005-02-14
; PRIOR APPLICATION NUMBER: 60/544,190
; PRIOR FILING DATE: 2004-02-13
; NUMBER OF SEQ ID NOS: 119966
; SEQ ID NO 97300
; LENGTH: 457
; TYPE: prt
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: (1)..(457)
; OTHER INFORMATION: Ceres Seq. ID no. 12736954
US-11-056-355B-97300

```

```
Query Match      8.1%; Score 80.5; DB 7; Length 457;
Best Local Similarity 29.1%; Pred. No. 6.4;
Matches 37; Conservative 15; Mismatches 46; Indels 29; Gaps 8;

QY 60 CEKQFYGNLKAATGTC--CVAMSYHVALIVSYKNAAGVTEKRIIDPSLFGSGPVTDTAM 117
DB 336 CMLOSLIDPTKVTGLGNCESQPSWSTYPQNTLTLDKSLCLE-----STGP----- 380
QY 118 RNACV---NTSCGSASVSSYAN--TAGNVY--RSPNSYLYD---NNLINTNCVLTKFS 167
DB 381 -NAPVALSTSCSPMLSEMETISASNNMLAAKSTNNSLCLVDDETNNLMASCKCVK-G 438

QY 168 LLSGCSGP 174
DB 439 EDSSCDP 445

RESULT 13
US-11-056-355B-97299
; Sequence 97299, Application US/11056355B
; Publication No. US20060150283A1
; GENERAL INFORMATION:
; APPLICANT: Brover, Vyacheslav
; APPLICANT: Alexandrov, Nickolai
; TITLE OF INVENTION: Sequence Determined DNA Fragments and Corresponding
; FILE REFERENCE: 2750-1590PUS2
; CURRENT APPLICATION NUMBER: US/11/056,355B
; PRIOR FILING DATE: 2005-02-14
; PRIOR APPLICATION NUMBER: 60/544,190
; PRIOR FILING DATE: 2004-02-13
; NUMBER OF SEQ ID NOS: 119966
; SEQ ID NO 97299
; LENGTH: 508
; TYPE: prt
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: (1)..(508)
; OTHER INFORMATION: Ceres Seq. ID no. 12736953
US-11-056-355B-97299

Query Match      8.1%; Score 80.5; DB 7; Length 508;
Best Local Similarity 29.1%; Pred. No. 7.3;
Matches 37; Conservative 15; Mismatches 46; Indels 29; Gaps 8;

QY 60 CEKQFYGNLKAATGTC--CVAMSYHVALIVSYKNAAGVTEKRIIDPSLFGSGPVTDTAM 117
DB 387 CMLOSLIDPTKVTGLGNCESQPSWSTYPQNTLTLDKSLCLE-----STGP----- 431
QY 118 RNACV---NTSCGSASVSSYAN--TAGNVY--RSPNSYLYD---NNLINTNCVLTKFS 167
DB 432 -NAPVALSTSCSPMLSEMETISASNNMLAAKSTNNSLCLVDDETNNLMASCKCVK-G 489

QY 168 LLSGCSGP 174
DB 490 EDSSCDP 496

RESULT 14
US-11-174-307B-1872
; Sequence 1872, Application US/11174307B
; Publication No. US20060143729A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai
; APPLICANT: Brover, Vyacheslav
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 2750-1601PUS2
; CURRENT APPLICATION NUMBER: US/11/174,307B
; PRIOR FILING DATE: 2005-06-30
; PRIOR APPLICATION NUMBER: 60/583,671
; PRIOR FILING DATE: 2004-06-30
```

```
; PRIOR APPLICATION NUMBER: 60/583,781
; PRIOR FILING DATE: 2004-06-30
; PRIOR APPLICATION NUMBER: 60/583,651
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 5544
; SEQ ID NO 1872
; LENGTH: 1182
; TYPE: prt
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: Pfam Name: peroxidase, Pfam Description: Peroxidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 21593467; NR Description: peroxidase ATR13a
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 31745133; NR Description: putative peroxidase
; OTHER INFORMATION: [Arabidopsis thaliana] >gi|9759059|dbj|BA809581.1| peroxidase
; OTHER INFORMATION: [Arabidopsis thaliana] >gi|1429219|emb|CAA67312.1| peroxidase
; OTHER INFORMATION: ATR13a [Arabidopsis thaliana] >gi|1402900|emb|CAA6966.1|
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 21593692; NR Description: putative peroxidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 6091756; NR Description: putative peroxidase
; OTHER INFORMATION: [Arabidopsis thaliana] >gi|25453221|sp|Q9S671|PE28 ARA7H Peroxidase
; OTHER INFORMATION: precursor (Arperox P28) (ATP39) >gi|15228606|ref|NP_187017.1|
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION:
; OTHER INFORMATION: GI Number: 51970002; NR Description: putative peroxidase
US-11-174-307B-1872

Query Match      8.1%; Score 80; DB 7; Length 1182;
Best Local Similarity 20.7%; Pred. No. 22;
Matches 34; Conservative 14; Mismatches 62; Indels 54; Gaps 7;

QY 23 CGTS--TASSPCTITRYVYDGCYARAHKMRQILMANGDCEKQFYGNLKAATGTC--CVA 79
DB 566 CGTTCCTTGTG-----CAATA-----TCATATTGGATCATGTCACAG 604
QY 80 WSYHVALIVSYKNAAGVTEKRIIDPSLFGSGPVTDTAMRNACVNTSCGSASVSSYANTAG 139
DB 605 CTCCTCTTGTGAAGAA-----TGATTTTAA--TGATCTCATCATGATGCTAG 652

QY 140 NVYRSPNSNYLYDNNLINTNCVLTKFSLLSGCSGPSAPDVSSC 183
DB 653 -----TTGCTCTTGAGGCTGCATCAATAAAC 678

RESULT 15
US-11-289-102-290
; Sequence 290, Application US/11289102
; Publication No. US20060121511A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Hyerin
; APPLICANT: Shaw, Peter M.
; APPLICANT: Clark, Edwin
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO
; FILE REFERENCE: 10338 NP
; CURRENT APPLICATION NUMBER: US/11/289,102
; PRIOR FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US 60/631,993
; PRIOR FILING DATE: 2004-11-30
; NUMBER OF SEQ ID NOS: 395
```

```

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 290
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-289-102-290

```

```

Query Match      8.0%; Score 79.5; DB 7; Length 723;
Best Local Similarity 22.0%; Pred. No. 14;
Matches 35; Conservative 26; Mismatches 71; Indels 27; Gaps 5;

```

```

QY 40 DGCYARAHKMRQILMNNNGYDCERKQFYGNLKASTGTCCVAMSYYHVALIVSYKNAAGVTEK 99
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 DACFSKFEHTNNLTWNCENC-GSYCY-----SSGPGCGQKXVFSSTSVTAHKQNSAQIP 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 100 RIIDPSLFSSGPTDT-----AMRNACVNTSCGSASVSYANTAGNVY 142
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 273 YALGKSLRPSAEWIEFTNDGKTELFCSINCLSAAYRKTVTSSGVQVSCHSCKTSAIPQY 332
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 143 YRSPSNSLYDNNLINTNCVL---TKESLLSGCSPSPAP 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 333 HLAMNGTLY--SFCSSCVVAFQNVFSKPKGTNSSAVP 369
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: September 16, 2006, 06:39:38  
Job time : 33 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2006, 06:18:35 ; Search time 40 Seconds  
(without alignments)  
445.003 Million cell updates/sec

Title: US-10-815-774-6

Perfect score: 991  
Sequence: 1 LASVIPDVATLNSLFNQIKN.....FSLHSGCSPSPAPADVSSCGF 185

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96.5	9.7	534	2 C69146	hypothetical prote
2	92	9.3	551	2 A83232	conserved hypothet
3	92	9.3	2395	1 S50820	surface protein ly
4	91	9.2	594	2 T50765	adhesion of calyx
5	90	9.1	243	2 T28802	hypothetical prote
6	89.5	9.0	927	2 T21772	hypothetical prote
7	89	9.0	1895	2 T15881	hypothetical prote
8	88.5	8.9	283	2 G69212	conserved hypothet
9	88.5	8.9	1251	2 T21389	hypothetical prote
10	85.5	8.6	534	2 T39903	serine-rich protei
11	85.5	8.6	1082	2 S37837	suppressor protein
12	85	8.6	594	2 T50764	adhesion of calyx
13	83.5	8.4	712	2 T33028	hypothetical prote
14	82.5	8.3	5376	2 T42215	zonadhesin - mouse
15	82	8.3	703	2 S67277	probable membrane
16	82	8.3	928	2 B72077	polymorphic membra
17	82	8.3	928	2 E86546	polymorphic outer
18	81.5	8.2	361	2 T19647	hypothetical prote
19	81	8.2	485	2 T49237	hypothetical prote
20	79.5	8.0	207	2 T22436	hypothetical prote
21	79.5	8.0	1894	2 UC4980	plexin 1 precursor
22	79	8.0	221	2 A59040	cysteine proteinas
23	79	8.0	275	2 T43004	hypothetical prote
24	79	8.0	732	2 B69749	hypothetical prote
25	78.5	7.9	725	2 T33174	band-6-protein - h
26	78.5	7.9	1067	2 T33174	probable zinc prot
27	77.5	7.8	422	2 T39306	mitogen-activated
28	77	7.8	814	2 T02195	hypothetical prote
29	77	7.8	1237	2 D71850	probable outer mem

30	77	7.8	2090	2 T30075	hypothetical prote
31	77	7.8	2153	2 T30074	hypothetical prote
32	77	7.8	2718	2 A23475	G surface protein
33	76.5	7.7	499	2 S39113	alpha-N-arabinofur
34	76.5	7.7	505	2 T28859	inward rectifier P
35	76.5	7.7	1234	2 C84465	hypothetical prote
36	76	7.7	735	2 A69146	hypothetical prote
37	76	7.7	934	2 T02742	probable ligand-ga
38	76	7.7	2704	2 S09118	G surface protein
39	75.5	7.6	594	2 T31744	hypothetical prote
40	75.5	7.6	1154	2 A86318	protein F15H18.11
41	75	7.6	443	2 F96990	MAD-specific glut
42	74.5	7.5	439	2 A36385	surface antigen se
43	74.5	7.5	980	2 T30089	probable zinc prot
44	74.5	7.5	1046	2 A26838	prestalk protein p
45	74.5	7.5	1335	2 T17508	glycoprotein Vp260

## ALIGNMENTS

## RESULT 1

C69146 hypothetical protein MTH359 - Methanobacterium thermoautotrophicum (strain Delta H)

C/Species: Methanobacterium thermoautotrophicum

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C/Accession: C69146

R/Smith, D.R.; Doucette-Stramm, L.A.; DeLonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;

Oliver, D.; Spadefora, R.; Vicaire, R.; Wang, Y.; Mierzbowski, J.; Gibson, R.; Ukwani, N

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, F.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funci

A/Reference number: A69000; PMID:98037514; PMID:9371463

A/Accession: C69146

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residue: 1-534 <MTH>

A/Cross-references: UNIPROT:Q26459; UNIPARC:UP1000006648A; GB:AE000821; GB:AE000666; NII

A/Experimental source: strain Delta H

C/Genetic:

A/Gene: MTH359

Query Match 9.7%; Score 96.5; DB 2; Length 534;  
Best Local Similarity 24.1%; Pred. No. 0.75;  
Matches 39; Conservative 23; Mismatches 71; Indels 29; Gaps 4;

QY	5	IPDVATLNSLFNQIKNQS-----CGTSTASSPCITFRYPV-----GCY 43
DB	358	LPNTVTVTSLASSLNKRPNDPYRGSTARYLASSACPDSPDIRSLASEITRGITSTF 417
QY	44	AAAHK-----MROILMNGVDEKQFVYGNLKASTGCCVAMGYHVAIVSYKASGVTEK 99
DB	418	SRBAVFGWNRNINIVSYFYNTKYGAVGTLKRTGCV---DHTHLVALRAAGIPAR 473
QY	100	RIIDPSLFSSGPTVDTAMRNACVNTSCGSASVSYANTAGNV 141
DB	474	VYHGTCNFTSGNVYGHVMAQLLVDPITWYAADATSSNSLSGV 515

## RESULT 2

A83232 conserved hypothetical protein PA3310 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004

C/Accession: A83232

R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Larbig, K.; Lam,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

A/Reference number: A82950; PMID:20437337; PMID:10984043

A/Accession: A83232

A/Status: preliminary



Qy 124 TSCGSASVSSYANTAGNVTYRSPNSY-LYDNNL-INTNCV 162  
 Db 143 GVC---VAQGTSYNNAMYSSTSTGNYCTYDNTCGLANQCV 179

## RESULT 6

T21772

hypothetical protein T01D3.6 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T21772; T24296

R:Steward, C.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19470

A:Accession: T21772

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 &lt;MIL&gt;

A:Cross-references: UNIPROT:O02364; UNIPARC:UPI0000086677; EMBL:Z81527; PIDN:CAB04279.1;

A:Experimental source: clone F35E12

R:Steward, C.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19870

A:Accession: T24296

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 &lt;M12&gt;

A:Cross-references: UNIPARC:UPI0000086677; EMBL:Z81110; PIDN:CAB03262.1; GSPDB:GN000023;

C:Genetics:

A:Gene: CESP:T01D3.6

A:Map position: 5

A:introns: 38/2; 87/3; 128/2; 242/2; 374/3; 416/3; 734/3; 842/3; 857/1

Query Match 9.0%; Score 89.5; DB 2; Length 927;

Best Local Similarity 20.1%; Pred. No. 5.9;

Matches 61; Conservative 25; Mismatches 68; Indels 149; Gaps 16;

Qy 4 VIPDVATLNSLFNOI--KNQSGTSTASPCITFRYPVGVGCVARAHKMQI-----LMN 55

Db 130 VEPDVASGFGCQNEIKCKDNKNSGKNA--DCYVANHQLNCLCKPGYTARK 176

Qy 56 NGYDECKQ-----FVYGNLIKAS-TGTCCVAMS----- 81

Db 177 NQRDDCMKQKQACMSGDPHYVYVYDGLRPFYQGTCPVFSQPCPTLLPAPYLMVSARAKNEL 236

Qy 82 ----YHVA---ILVSYKN-----ASGVTEKRITIDPSLFS-----S 109

Db 237 PGKGYHISQVSEVEVDLHMLTIHVDRSKTALVNGV--QVLTPLYFPKNTKMTVTVRVS 293

Qy 110 GP-----VTDTAWRNACV-----NTSCGSA-----SVSSYANTAGNV 141

Db 294 GSTFTTENDQGVVVFITTYNSLCVOVPDIPEFNGATTLGLAGNIDGKKLDVYVKNKNSV 353

Qy 142 Y-----YRSPSN-----SYLYDNNLI-----NTNCVLTRESLFS 170

Db 354 LAIKSRQENNNHADMFKETEDTWTITDKFLIIRPGQENCINGQTLDDNNTNCVSTISLAQ 413

Qy 171 GCS 173

Db 414 SCA 416

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

A:Accession: T15881  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1895 <PAU>  
 A:Cross-references: UNIPARC:UPI000004CAE8; EMBL:U00065; NID:g495681; PID:g495684; PIDN:?

## RESULT 7

T15881

hypothetical protein D1044.3 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T15881

R:Pauley, A.

submitted to the EMBL Data Library, June 1994

A:Reference number: Z18423

A:Accession: T18423

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 &lt;MIL&gt;

A:Cross-references: UNIPROT:O02364; UNIPARC:UPI0000086677; EMBL:Z81527; PIDN:CAB04279.1;

A:Experimental source: clone F35E12

R:Steward, C.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19870

A:Accession: T24296

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-927 &lt;M12&gt;

A:Cross-references: UNIPARC:UPI0000086677; EMBL:Z81110; PIDN:CAB03262.1; GSPDB:GN000023;

C:Genetics:

A:Gene: CESP:T01D3.6

A:Map position: 5

A:introns: 38/2; 87/3; 128/2; 242/2; 374/3; 416/3; 734/3; 842/3; 857/1

Query Match 9.0%; Score 89.5; DB 2; Length 927;

Best Local Similarity 20.1%; Pred. No. 5.9;

Matches 61; Conservative 25; Mismatches 68; Indels 149; Gaps 16;

Qy 4 VIPDVATLNSLFNOI--KNQSGTSTASPCITFRYPVGVGCVARAHKMQI-----LMN 55

Db 130 VEPDVASGFGCQNEIKCKDNKNSGKNA--DCYVANHQLNCLCKPGYTARK 176

Qy 56 NGYDECKQ-----FVYGNLIKAS-TGTCCVAMS----- 81

Db 177 NQRDDCMKQKQACMSGDPHYVYVYDGLRPFYQGTCPVFSQPCPTLLPAPYLMVSARAKNEL 236

Qy 82 ----YHVA---ILVSYKN-----ASGVTEKRITIDPSLFS-----S 109

Db 237 PGKGYHISQVSEVEVDLHMLTIHVDRSKTALVNGV--QVLTPLYFPKNTKMTVTVRVS 293

Qy 110 GP-----VTDTAWRNACV-----NTSCGSA-----SVSSYANTAGNV 141

Db 294 GSTFTTENDQGVVVFITTYNSLCVOVPDIPEFNGATTLGLAGNIDGKKLDVYVKNKNSV 353

Qy 142 Y-----YRSPSN-----SYLYDNNLI-----NTNCVLTRESLFS 170

Db 354 LAIKSRQENNNHADMFKETEDTWTITDKFLIIRPGQENCINGQTLDDNNTNCVSTISLAQ 413

Qy 171 GCS 173

Db 414 SCA 416

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

A:Reference number: Z19416  
 A:Accession: T21389  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1251 <WII>  
 A:Cross-references: UNIPROT:Q09550; UNIPARC:UPI000013BF4C; EMBL:Z47072; P1DN:CAA87369.1  
 A:Experimental source: clone F26C11  
 C:Genetics:  
 A:Gene: CESP:F26C11.3  
 A:Map position: 2  
 A:Introns: 24/1; 111/1; 208/1; 272/1; 380/1; 394/3; 485/3; 586/3; 630/3; 669/3; 713/3; 7

Query Match  
 Best Local Similarity 23.7%; Score 88.5; DB 2; Length 1251;  
 Matches 36; Conservative 24; Mismatches 55; Indels 37; Gaps 6;

QY 21 QSC-GTSTASSPCITFRYPVDCGYARAHKROILMNGDYCEKQFYVGNLAKSTGTCCVA 79  
 DB 986 QGCKATSTGTPT-TTNWPTGGT-TTTLDSGEIILSES----- 1021  
 QY 80 WSYHVALIVSYKNAAGVTEKRIIDPSLFSGPVTDAMNACVNTSCGSASVSYANTAG 139  
 DB 1022 -----LIAYKCTIVLMQILINPSKNTTRTET-TSDAGCKATSSGTTSTWSPGTGG 1073  
 QY 140 NVYRSPSPNSLYDNNLINTNCVLTKFSLISG 171  
 DB 1074 TTVSRFTNSNNPIDSTLET-----TTFAMPYG 1101

RESULT 10  
 T39903  
 A:Reference number: Z19416  
 A:Accession: T21389  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-534 <LVN>  
 A:Cross-references: UNIPROT:Q04317; UNIPARC:UPI0000069AEB; EMBL:AL033534; P1DN:CAA22127.  
 A:Experimental source: strain 972h-; cosmid c215  
 C:Genetics:  
 A:Gene: SPDB:SPBC215.13  
 A:Map position: 2  
 C:Superfamily: pig submaxillary mucin

Query Match  
 Best Local Similarity 24.2%; Score 85.5; DB 2; Length 534;  
 Matches 45; Conservative 32; Mismatches 86; Indels 23; Gaps 5;

QY 6 PDVATLNLFNQIKKQSCGTSTASSPCITFRYPVDCGYARAHKROILMNGDYCEKQFY 65  
 DB 78 PTSSSEPTIFSSSATPS-ETNSYSSPVSSYSDPATSS-----QLPSSSTSPSSSE 127  
 QY 66 YGNLKASTGTCCVAMSYHVALIVSYKNAAGVTEKRIIDPSLFSGPVTDAMNACVNTS 125  
 DB 128 YTPSSTESSLIDPSVSSAILPS-----STVEVASISSSLSSDPLTSTSPSSSSTS 183  
 QY 126 CGSASVSS-YANTAGNVYRSPNSLYDNNLI-----NTNVCVLTKEFSLISGCSPPS 176  
 DB 184 SSQPSVSTSSSTSPSSAAPTSTSSSTLSSSVSSSSSPSSSSTLSTSSSTSTSTST 243  
 QY 177 APDVSS 182  
 DB 244 SSSSSS 249

RESULT 11  
 S37837  
 A:Reference number: Z19416  
 A:Accession: T21389  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1251 <WII>  
 A:Cross-references: UNIPROT:Q09550; UNIPARC:UPI000013BF4C; EMBL:Z47072; P1DN:CAA87369.1  
 A:Experimental source: clone F26C11  
 C:Genetics:  
 A:Gene: CESP:F26C11.3  
 A:Map position: 2  
 A:Introns: 24/1; 111/1; 208/1; 272/1; 380/1; 394/3; 485/3; 586/3; 630/3; 669/3; 713/3; 7

N:Alternate names: protein YKL020C  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 31-Dec-2004  
 C:Accession: S37837; S40696  
 R:Rieger, M.  
 submitted to the Protein Sequence Database, March 1994  
 A:Reference number: S37832  
 A:Accession: S37837  
 A:Molecule type: DNA  
 A:Residues: 1-1082 <RIE>  
 A:Cross-references: UNIPROT:P35210; UNIPARC:UPI0000135D64; EMBL:Z28020; NID:9486010; PID  
 A:Experimental source: strain S288C  
 R:Burkett, T.J.; Garfinkel, D.J.  
 Yeast 10, 81-92, 1994  
 A:Title: Molecular characterization of the SPT23 gene: a dosage-dependent suppressor of  
 A:Reference number: S40695; MUID:94262317; PMID:8203154  
 A:Accession: S40696  
 A:Molecule type: DNA  
 A:Residues: 136-714, 'P', 716-737, 'ERKLYVLLATYGY' <BUR>  
 A:Cross-references: UNIPARC:UPI0000168E4A; EMBL:L24760; NID:9404815; P1DN:AAA20575.1; P1  
 F:742-774/Domain: ankyrin repeat homology <AN1>

Query Match  
 Best Local Similarity 21.7%; Score 85.5; DB 2; Length 1082;  
 Matches 40; Conservative 27; Mismatches 54; Indels 63; Gaps 8;

QY 27 TASSPCITFRYPVD-GCYARAHKROILMNGDYCEKQFYVGNLAKSTGTCCVAMSYHA 85  
 DB 259 SGNNSCINFDLPTRIVCYCRHKK-----ATNGF----- 288  
 QY 86 ILVSYKNAAGVTEKRIIDPSLFS-----SSGPVTDAMNACVNTSCGSASVSYANTAG 140  
 DB 289 IFLRDHNGDILAKITTPIMMDKKNASNTTPTSTISMAVSM--TNDTRSPSSPSQSD 346  
 QY 141 VYRS-----PSNSLYDNNLINTNCVL-----TKFSLISGCSPPSAP 178  
 DB 347 LMFPEFPLPSSNS---KNFVISTNCMLDSNCNNNNNDNDKNNIKITNYAMNNRHHFSP 403  
 QY 179 DVSS 182  
 DB 404 NSSS 407

RESULT 12  
 T50764  
 A:Reference number: Z19416  
 A:Accession: T21389  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-594 <ARA>  
 A:Cross-references: UNIPROT:Q95X23; UNIPARC:UPI000009C6A5; EMBL:AB027507; P1DN:BAA77842  
 A:Experimental source: strain Nossen  
 C:Genetics:  
 A:Gene: ACE  
 A:Introns: 25/1; 155/2; 259/1; 387/3; 577/2  
 C:Superfamily: alcohol oxidase

Query Match  
 Best Local Similarity 24.6%; Score 85; DB 2; Length 594;  
 Matches 31; Conservative 13; Mismatches 46; Indels 36; Gaps 5;

QY 25 TSTASSPCITFRY---PVD---GCYARAHKROILMNGDYCEKQFYVGNLAKST-- 73

```

Db      442 TIVDDNPSTVFENFKAPVLDQRCEAIRLWSKVYTSKREPLNYQCDKQNVHKKMLSLSYVA 501
QY      74 -----GTCVAMSYVVALIVS-----YKASGVTEKRIIDPSLPS 108
Db      502 NITLRKQINDTYSMAQFCQDITVVTIWHHGCLVGKVVSPNKKVLGVDRLRVIDDSTPD 561
QY      109 SGPVTD 114
Db      562 ESPGTN 567

RESULT 13
T33028
hypothetical protein T09A12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33028
R: Hawkins, J.; Fullon, B.; Gillam, B.
submitted to the EMBL Data Library, February 1998
A:Description: The sequence of C. elegans cosmid T09A12.
A:Reference number: 221265
A:Accession: T33028
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-712 <HAM>
A:Cross-references: UNIPARC:UPI000017B882; EMBL:AF047660; PIDN:AAC04434.1; GSPDB:GN00022
A:Experimental source: strain Bristol N2; clone T09A12
A:Gene: CSP:T09A12.4
A:Map position: 4
A:Introns: 32/1; 158/1; 220/1; 244/3; 295/1; 319/1; 436/3; 543/2; 634/3

Query Match: 8.4%; Score 83.5; DB 2; Length 712;
Best Local Similarity 21.5%; Pred. No. 15;
Matches 49; Conservative 20; Mismatches 66; Indels 93; Gaps 11;

QY      23 CGTSTASSPCTIFRYVDCGYA-RAHKMRQILMNNGYDEKQFVGNLK-ASTGTCCVAM 80
Db      251 CGTDSNG-----IHFGVDACACSAFPRRTIVLVANKDYSCNKGKCTVVVDGSAQKCRAC 305
QY      81 SYHVALIVSY-----KNASGVTEKRIIDPSL-----FSSGP 111
Db      306 RFRKCISSGMDKNKSVQHRDAIGKYSAGV--KRELSPDAEFEPBAKYSTVSEPTSSGSPS 363
QY      112 -----VTDAMRNACVNTSGSASVSS-----YANT 137
Db      364 GGNQNVSSPAGIPRVPSLTRITQASTCMMSCGQKSVLHELICRONFLTEOROLFYAGC 423
QY      138 AGNVYRSPNSLYLNNLINTNCVLTFFSLSGCSPSPAPDVSSCGF 185
Db      424 LGD-WERKPS--IENQTL-----SELTDFSSCMF 450

RESULT 14
T42215
zonadhesin - mouse
N:Alternate names: sperm-specific membrane protein
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42215
R: Gao, Z.; Garbers, D.L.
J. Biol. Chem. 273, 3415-3421, 1998
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane pro
A:Reference number: Z22080; M0ID:9813114; PMID:9452463
A:Accession: T42215
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5376 <GAO>
A:Cross-references: UNIPROT:O88799; UNIPARC:UPI000002A15F; EMBL:U97068; NID:93327420; PI
C:Gene: Zan
A:Map position: 5

```

```

C:Function:
A:Description: functions in multiple cell adhesion processes
A:Note: found exclusively on the apical region of the sperm head
C:Keywords: cell adhesion

Query Match: 8.3%; Score 82.5; DB 2; Length 5376;
Best Local Similarity 24.9%; Pred. No. 1.8e+02;
Matches 43; Conservative 12; Mismatches 53; Indels 65; Gaps 10;

QY      24 GTS-TASSPCTIFRYVDCGYARAHMRQILMNNGYDEKQFVGNLKA-STGTCVAMSGY 82
Db      4530 GTSFAPSTC-----EEGCV-----CEPDVYLSNDKCVPSSEC----- 4562
QY      83 HVALIVSYKASGVTEKRIIDPSLPSGPVTDPA--RNACVNTSGSASVSSYANTAGN 140
Db      4563 -----GCKDAIGV---LIPES-----KTWVSRCCTKNCCTCKGGTVQCHD----- 4598
QY      141 VYRSPNSLYLNNLINTNCV-----LTKFSLSGCSPSPAPDVSS-----CG 184
Db      4599 --FSCPTGSRCDNNEGNSNCVTYALKCPAHLTYNCLPSCLPSCSPDPLGCG 4649

RESULT 15
S67277
probable membrane protein YOR365C - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O6661
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C:Accession: S67277
R: Delius, H.; Hebling, U.; Hofmann, B.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67261
A:Accession: S67277
A:Molecule type: DNA
A:Residues: 1-703 <DEL>
A:Cross-references: UNIPROT:Q08844; UNIPARC:UPI000006A32E; EMBL:Z75272; NID:91420789; PI
A:Experimental source: strain S288C
A:Gene: MIPS_YOR365C
A:Cross-references: SGD:S0005892
A:Map position: 15R
A:Superfamily: probable membrane protein YAL053W
C:Keywords: transmembrane protein
F:2-18/Domain: transmembrane #status predicted <TM1>
F:208-224/Domain: transmembrane #status predicted <TM2>
F:248-264/Domain: transmembrane #status predicted <TM3>
F:379-395/Domain: transmembrane #status predicted <TM4>
F:431-447/Domain: transmembrane #status predicted <TM5>
F:454-470/Domain: transmembrane #status predicted <TM6>
F:546-562/Domain: transmembrane #status predicted <TM7>
F:575-591/Domain: transmembrane #status predicted <TM8>
F:599-615/Domain: transmembrane #status predicted <TM9>
F:616-632/Domain: transmembrane #status predicted <TM10>

Query Match: 8.3%; Score 82; DB 2; Length 703;
Best Local Similarity 19.0%; Pred. No. 20;
Matches 39; Conservative 34; Mismatches 66; Indels 66; Gaps 6;

QY      31 PCITRRYP-----VDGCVARAHMRQILMNNGYDEKQFVGN----- 68
Db      157 PSITVINIPDLAYIVTAVSTIDKEFTPLACVQWLSNGRVTQREYLSMNLVILLTSGI 216
QY      69 -----LKASNGTCVAMSYVVALIVSYKN-----ASGV 96
Db      217 MGSVYSLQGYTIVTIRLASISILLFQNLAILAMISVSLPPIVAANTONFQWSKCI 276
QY      97 TE-----KRIIDPSLPSGPVTDPAWNAVC-----NTSCGSASVSYANTAGNYRSP 146
Db      277 IININMQRLFDWYVAATSSPTVYVRNKEVLSISVQKSLNSKIISASGNLNG--IBSSQ 334
QY      147 SNSLYLNNLINTNCVLTFFSLSG 171
Db      335 KNDLLYTSNLRNSNDYLSKILVLRG 359

```

Mon Sep 18 16:01:20 2006

us-10-815-774-6.rpr

Page 6

Search completed: September 16, 2006, 06:24:04  
JOB time : 43 secs



```
Q6MGS8_BDEBA
ID Q6MGS8_BDEBA PRELIMINARY; PRT; 252 AA.
AC Q6MGS8;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DE 07-FEB-2006, entry version 9.
DI Hypothetical protein precursor.
GN OrderedLocustNames=Bd3844;
OS Bdellovibrio bacteriovorus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibrio.
NCBI_TaxID=959;
OX [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RC PubMed=14752164; DOI=10.1126/science.1093027;
RX Rendicic S., Jędrap P., Rosinus A., Eppinger M., Baar C.,
RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
RA Sockeet R.E., Schuster S.C.,
RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
RT genomic perspective.";
RL Science 303:689-692(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
CC EMBL; BX842656; CAB81201.1; -; Genomic_DNA.
CC
CC DR BioCyc; BBAC264462; BD3844-MONOMER; -.
CC Complete proteome; Hypothetical protein; Signal.
CC STGNLU 1 Potential.
CC SEQUENCE 252 AA; 26633 MW; 88B0A8661F93118C CRC64;
SQ
Query Match 14.4%; Score 142.5; DB 2; Length 252;
Best Local Similarity 32.3%; Pred. No. 4e-05;
Matches 40; Conservative 17; Mismatches 54; Indels 13; Gaps 4;
QY 5 IPDVATLSLNFQIK-----NQSGCTSYASSPCITFRYPVDCYARAHKMQIIMNNGYDC 60
DB 66 IPDVGSYADLEREKYVRDTRFTADAPFARRLTWMTPDDCCARALATLVEHQFTT 125
QY 61 EKO-FVYGNLKAATGTC---CVAMSYHVALIVSYKNSGVTEKRIIDPSLFSSGPTVDTA 116
DB 126 PKKIFVGNLYAPFQNAFGSGVSWWYHVAVTYRVGN-----EYVFPDAINPEKPTLVE 180
QY 117 WRNA 120
DB 181 WNKA 184
RESULT 3
ID Q5WM27_LEGPH PRELIMINARY; PRT; 274 AA.
AC Q5WM27;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DE 07-FEB-2006, entry version 8.
DI Hypothetical protein.
GN OrderedLocustNames=lp11633;
OS Legionella pneumophila (strain Lens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae; Legionella.
NCBI_TaxID=297245;
OX [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP PubMed=15467720; DOI=10.1038/ng1447;
RX Cazalec C., Rusniok C., Brueggemann H., Zidane N., Magnier A., Ma L.,
RA Tichit M., Jarraud S., Bouchier C., Vandenesch F., Kunst F.,
RA Etienne J., Glaser P., Buchrieser C.,
RT "Evidence in the Legionella pneumophila genome for exploitation of
RT host cell functions and high genome plasticity.";
RL Nat. Genet. 36:1165-1173(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
```

```
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
CC EMBL; CR628337; CAH15873.1; -; Genomic_DNA.
CC
CC DR LegiList; lp11633; -.
CC Complete proteome; Hypothetical protein.
CC SEQUENCE 274 AA; 31408 MW; 2E87A5626A211293 CRC64;
SQ
Query Match 12.9%; Score 127.5; DB 2; Length 274;
Best Local Similarity 31.9%; Pred. No. 0.0012;
Matches 44; Conservative 23; Mismatches 35; Indels 37; Gaps 8;
QY 3 SVIPDVAT--LNSLFNQIKNSCGTSTASSP-----CITFRYPVDCYARAHKMQI 52
DB 72 STVPKVSYSYELAMFNILRD-----TRFLYSPDPDPQRRISWLYPDDGCFARALSRIK 127
QY 53 LMNNGYDCE-----KQFVGNLK-----ASTGTCVAMSYHVALIVSYKNSGVTEKRI 101
DB 128 L-----DSEHFVIPAKTFVFGDLEMQTPYSSEGT--VSMWYHVASVANNYKGTI-----YV 175
QY 102 IDPSLFSSGPTVDTAWRN 119
DB 176 LDPAAKPEGPMILDMVN 193
```

```
RESULT 4
ID Q5ZUX4_LEGPH PRELIMINARY; PRT; 294 AA.
AC Q5ZUX4;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DE 07-FEB-2006, entry version 7.
DI Hypothetical protein.
GN OrderedLocustNames=lp91668;
OS Legionella pneumophila subsp. pneumophila (strain Philadelphia 1 /
OS ATCC 33152).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Legionellaceae; Legionella.
NCBI_TaxID=272624;
OX [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP PubMed=15448271; DOI=10.1126/science.1099776;
RX Chien M., Morozova I., Shi S., Sheng H., Chen J., Gomez S.M.,
RA Asanami G., Hill K., Nuara J., Feder M., Rineer J., Greenberg J.J.,
RA Stesenko V., Park S.H., Zhao B., Teplickaya E., Edwards J.R.,
RA Pamou S., Georgioulou A., Chou I.-C., Iannuccilli W., Uiz M.E.,
RA Kim D.H., Geringer-Sameth A., Goldsberry C., Morozov P., Fischer S.G.,
RA Segal G., Qu X., Rzhetsky A., Zhang P., Cayanis E., De Jong P.J.,
RA Du J., Kalachikov S., Shuman H.A., Russo J.J.,
RT "The genomic sequence of the accidental pathogen Legionella
RT pneumophila.";
RL Science 305:1966-1968(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
CC EMBL; AE017354; AAU27748.1; -; Genomic_DNA.
CC Complete proteome; Hypothetical protein.
CC SEQUENCE 294 AA; 33731 MW; 8850B3A8D18669CD CRC64;
SQ
Query Match 12.9%; Score 127.5; DB 2; Length 294;
Best Local Similarity 31.2%; Pred. No. 0.0013;
Matches 43; Conservative 23; Mismatches 35; Indels 37; Gaps 8;
QY 3 SVIPDVAT--LNSLFNQIKNSCGTSTASSP-----CITFRYPVDCYARAHKMQI 52
DB 92 STVPKVSYSYELAMFNILRD-----TRFLYSPDPDPQRRISWLYPDDGCFARALSRIK 147
QY 53 LMNNGYDCE-----KQFVGNLK-----ASTGTCVAMSYHVALIVSYKNSGVTEKRI 101
DB 148 L-----DSEHFVIPAKTFVFGDLEMQTPYSSEGT--VSMWYHVASVANNYKGTI-----YV 195
QY 102 IDPSLFSSGPTVDTAWRN 119
```



DB 196 LDPAAKEGPMILIDWYN 213

RESULT 5

OSX4N7\_LEGPA PRELIMINARY; PRT; 274 AA.

AC 05X4N7; LEGPA

DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.

DT 23-NOV-2004, sequence version 1.

DT 07-FEB-2006, entry version 8.

DE Hypothetical protein.

OS OrderedLocNames=lpri639;

GN Legionella pneumophila (strain Paris).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;

OC Legionellaceae; Legionella.

OX NCBI\_Taxid=297246;

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RX PubMed=15467720; DOI=10.1038/ng1447;

RA Cazalet C., Rusniok C., Brueggemann H., Zidane N., Magnier A., Ma L.,

RA Tichtit M., Jarrad P., Bouchier C., Vandenesch F., Kunst F.,

RA Etienne J., Glaser P., Buchrieser C.;

RT "Evidence in the Legionella pneumophila genome for exploitation of

RT host cell functions and high genome plasticity.";

RT Nat. Genet. 36:1165-1173(2004).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonDerivs license

CC -----

DR EMBL; CR628336; CAH12791.1; -; Genomic\_DNA.

DR LegioLst; lpr1639; -

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 274 AA; 31284 MW; 5B302361FELC834B CRC64;

Query Match 12.7%; Score 125.5; DB 2; Length 274;

Best Local Similarity 31.2%; Pred. No. 0.0018;

Matches 43; Conservative 22; Mismatches 36; Indels 37; Gaps 8;

QY 3 SVTPDVAT---LNSLPNOKKSCGTSASP-----CIFRYPVDCGYARAHMQR1 52

DB 72 STVPKVSSTYDELNMNMLIRD---TRFLYSPDKPDRRISMLVPDDCFAPAAISR1K 127

QY 53 LMANGYDCE-----KQFYGNLK-----ASTGTCVANSYHVALIVSYKNSGVTEKRI 101

DB 128 L-----DSHFVTPAKIFAFGIDEMQTPSSSEST--VSMWYHVASAVVNTGRT-----YV 175

QY 102 IPPSLFSSGQVTDTRNRN 119

DB 176 LDPAAKEGPMILIDWYN 193

RESULT 6

Q6MUD7\_BDEBA PRELIMINARY; PRT; 267 AA.

AC Q6MUD7; BDEBA

DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.

DT 05-JUL-2004, sequence version 1.

DT 07-FEB-2006, entry version 9.

DE Hypothetical protein.

OS OrderedLocNames=Bd2842;

GN Bdellovibrio bacteriovorus.

OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;

OC Bdellovibrionaceae; Bdellovibrio.

OX NCBI\_Taxid=959;

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RX STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;

RX PubMed=14752164; DOI=10.1126/science.1093027;

RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,

RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,

RA Sockett R.E., Schuster S.C.;

RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a

RT genomic perspective.";

RL Science 303:689-692(2004).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonDerivs license

CC -----

DR EMBL; BX842653; CAE80623.1; -; Genomic\_DNA.

DR BioCyc; BBAC264462; BD2842-MONOMER; -

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 267 AA; 29959 MW; 9AD3258D47989064 CRC64;

Query Match 11.3%; Score 111.5; DB 2; Length 267;

Best Local Similarity 26.3%; Pred. No. 0.038;

Matches 45; Conservative 28; Mismatches 63; Indels 35; Gaps 9;

QY 2 ASVIDVATLNS-----LFNQIKKSCGTSASPCTFRYPVDCGYARAHMQR11MN 55

DB 68 ATLQGPXKATWSSEADAKLFSLSKSG-----DIPFERSLNGCERRAHMSRLML 118

QY 56 NGYDEKQFYGN-----LKASTGT---CCVANSYHVALIVSYKNSGVTEKRI1DPSL 106

DB 119 KGITPLKVPASVNEDESRLRPNTKGMVTDMKHYAPVLYKKSGLV--PYVMDPSL 177

QY 107 FSSG--PVTDTMR-----NACVNTSCGASVSYANTAGNVYRSPNSY 150

DB 178 EKKAVPVS--WQATMTRHNPMMKYNLKTTPATY--NDAGITRVNFKONDF 225

RESULT 7

ID 026459\_METTH PRELIMINARY; PRT; 534 AA.

AC 026459; METTH

DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.

DT 01-JAN-1998, sequence version 1.

DT 07-FEB-2006, entry version 23.

DE Conserved protein.

GN OrderedLocNames=MTN359; ORFNames=MTN\_359;

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.

OX NCBI\_Taxid=187420;

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RX STRAIN=Delta H;

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Delouhery C., Lee H.-M., Dubois J.,

RA Altedge T., Bashirzaden R., Blakey J.D., Cook R., Gilbert K.,

RA Harrison D., Hoang L., Keagle P., Lunn W., Pothier B., Qiu D.,

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,

RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT deltaH: functional analysis and comparative genomics.";

RT J. Bacteriol. 179:7135-7155(1997).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NonDerivs license

CC -----

DR EMBL; AE000666; AAB84865.1; -; Genomic\_DNA.

DR PIR; C69146; C69146.

DR BioCyc; MTH187420:MTN359-MONOMER; -

DR InterPro; IPR002931; TmnsGluase\_1like.

DR Pfam; PF01841; TmnsGlu\_core; 1.

DR SMART; SM00460; TGC; 1.

KW Complete proteome.

SQ SEQUENCE 534 AA; 57628 MW; 588412BD8D342E26 CRC64;

Query Match 9.7%; Score 96.5; DB 2; Length 534;

Best Local Similarity 24.1%; Pred. No. 2.4; Indels 29; Gaps 4;

Matches 39; Conservative 23; Mismatches 71; Indels 29; Gaps 4;

QY 5 IPDVATLNSLPNOKKSCGTSASPCTFRYPVD-----GCY 43

RESULT 9	
Q9HYT4_PSEAE	
ID Q9HYT4_PSEAE	PRELIMINARY; PRT; 551 AA.

```

RESULT 10
Q27167 PARTE
ID Q27167 PARTE PRELIMINARY; PRT; 2395 AA.
AC Q27167;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE 51B type surface protein (Surface antigen 51B) .
OS Paramoecium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramoecidae; Paramoecium.
OX NCBI_TaxID=5688;
RN [1]

```

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Stock 51;  
 RX MEDLINE=95098630; PubMed=7800503;  
 RA Scott J., Leeck C., Forney J.;  
 RT "Analysis of the micronuclear B type surface protein gene in  
 RL Paramoecium tetraurelia.";  
 RN Nucleic Acids Res. 22:5079-5084 (1994).  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Stock 51;  
 RA Matuda A., Forney J.D.;  
 RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Stock 51;  
 RA Scott J.M., Leeck C.L., Forney J.D.;  
 RT "Molecular and genetic analyses of the B Type surface protein gene  
 RL from Paramoecium tetraurelia.";  
 RN Genetics 133:189-198 (1993).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 CC EMBL: U07603; AAA81947.1; -; Genomic\_DNA.  
 DR EMBL: U04795; AAA6710.1; -; Unassigned\_DNA.  
 DR PIR: S50820; S50820.  
 DR HSP: P83469; IMP2.  
 DR GO: GO:0005576; C:extracellular region; IEA.  
 DR GO: GO:0006030; P:chitin binding; IEA.  
 DR GO: GO:0006030; P:chitin metabolism; IEA.  
 DR GO: GO:0007275; P:development; IEA.  
 DR InterPro: IPR002895; Paramoecium\_SA.  
 DR InterPro: IPR002557; Para\_Chitin\_bd.  
 DR InterPro: IPR003659; Plexin-like.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF01508; Paramoecium\_SA; 31.  
 DR SMART: SM00639; PSA; 26.  
 DR SMART: SM00423; PSI; 9.  
 DR PROSITE: PS00940; CHIT\_BIND\_II; 3.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 SQ SEQUENCE 2395 AA; 246433 MW; 8632E022649B2D61 CRC64;  
 Query Match 9.3%; Score 92; DB 2; Length 2395;  
 Best Local Similarity 22.4%; Pred. No. 38;  
 Matches 57; Conservative 22; Mismatches 76; Indels 100; Gaps 12;  
 QY 10 TNSLFFNQIKNOSGCTSTASPCITRYPD-----GCTARAHKROILMNG 57  
 DB 487 TVNSTVAGCVDTKCNLSLAQITC-----DKDLNKKACIMKGCYKRECVLASSTTAT 539  
 QY 58 YDCENQFVYGNLKASGTGCV-----AMSYHVALVSYKASGV----- 96  
 DB 540 AACCQYVDV-GCTLSNLTGCVPLKCEALITTEACNINRLQVTSVYKSYPLCGWNGSSCI 598  
 QY 97 -----TEKRIIDPSLFSG-----PYTDTA----- 116  
 DB 599 DKAQSTAPRTTATTSOCGYKSCVANNPVNGSIQCCDPLPTTCARKESTENCEITRTGF 658  
 QY 117 ---WF---NACVNTSCGASVSVYNTAGNVTYRSPNSVLYDNMLINTNCVLTKESLL 169  
 DB 659 PCTLMNATSACTYKSCSTASVTT---TTGFLTFVSTNCLAV---LSNSACIAN--NTA 710  
 QY 170 SGCSPPAPDVSSCG 184  
 DB 711 DGCIPKP-----SSCG 721

RESULT 11  
 OS6X03\_ARATH PRELIMINARY; PRT; 294 AA.  
 AC OS6X03;  
 DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.  
 DT 10-MAY-2005, sequence version 1.

DT 07-FEB-2006, entry version 3.  
 DE Hypothetical protein At1g72970.  
 GN Name=At1g72970;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Totoki Y., Seki M., Ishida J., Nakajima M., Enju A., Kamiya A.,  
 RA Narusaka M., Shin-I T., Nakagawa M., Sakamoto N., Oishi K., Kohara Y.,  
 RA Kobayashi M., Toyoda A., Sakaki Y., Sakurai T., Iida K., Akiyama K.,  
 RA Satou M., Toyoda T., Konagaya A., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinzaki K.;  
 RT "Large-scale analysis of RIKEN Arabidopsis full-length (RAFL) cDNAs";  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonDerivs License  
 CC -----  
 CC EMBL: AK21876; BAD94191.1; -; mRNA.  
 DR GO: GO:0050660; P:AD binding; IEA.  
 DR GO: GO:0016164; P:oxidoreductase activity, acting on CH-OH gr. . .; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR InterPro: IPR001472; GMC\_oxred.  
 DR InterPro: IPR007667; GMC\_oxred\_C.  
 DR Pfam: PF05199; GMC\_oxred\_C; 1.  
 DR Pfam: PF00732; GMC\_oxred\_N; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 294 AA; 32889 MW; AD9E633777119621 CRC64;  
 Query Match 9.2%; Score 91; DB 2; Length 294;  
 Best Local Similarity 24.6%; Pred. No. 3.9;  
 Matches 31; Conservative 14; Mismatches 45; Indels 36; Gaps 5;  
 QY 25 TSTASPCITERY--PVD--GCTARAHKROILMNGY---DCEKQFVYGNLKAFT-- 73  
 DB 142 TVNDNPSVTFYFPHVPDLQKCEALRLSVKSVYNSNPLNTYQCDQVHMLSLSYVA 201  
 QY 74 -----GTCVAMSYHVALVY-----YKMSGTEKRIIDPSLFS 108  
 DB 202 NINLRPKQINDTKNSAQCCKDTVTVTIMYHGCLVGVKVSPPRKVLGVDRRLVIGDSTPD 261  
 QY 109 GGPVTD 114  
 DB 262 ESEPTN 267

RESULT 12  
 HTH\_ARATH STANDARD; \*PRT; 594 AA.  
 AC Q9S746; Q9S6X23;  
 DT 26-APR-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-MAY-2000, sequence version 1.  
 DT 07-MAR-2006, entry version 31.  
 DE Protein HTH-HEAD precursor (Protein ADHESION OF CALYX EDGES).  
 GN Name=HTH; Synonyms=ACE; OrderedLocustNames=At1g72970;  
 GN ORFNames=F3N23.17;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND VARIANT LYS-478.  
 RC STRAIN=cv. Columbia, and cv. No-0;  
 RA Nakatani-Goto M., Arai T., Iwabuchi M.;  
 RT "ADHESION OF CALYX EDGES, a gene involved in the regulation of  
 RL Plant Cell Physiol. 39:864-864 (1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], MOTAGENESIS OF GLY-218; ARG-227;



```
RESULT 13
O1804 CAEEL PRELIMINARY; PRT; 243 AA.
ID O1804 CAEEL
AC O1804
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DE Hypothetical protein C53B7.3.
GN ORFName=C53B7.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
   investigating biology.";
RL Science 282:2012-2018(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs license
CC -----
DR EMBL, U42830; AAC48274.1; -; Genomic_DNA.
DR PIR, T28802; T28802.
DR HSSP, P02876; 9MGA.
DR Ensembl, C53B7.3; Caenorhabditis elegans.
DR Wormbase, WBGene0016894; C53B7.3.
DR WormPep, C53B7.3; CE06973.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 243 AA; 26299 MW; 6C1E33CB1CA2D16B CRC64;

Query Match
Best Local Similarity 26.7%; Pred. No. 3.9;
Matches 43; Conservative 14; Mismatches 64; Indels 40; Gaps 9;

OY 22 SGGTSSASPCITFRPVGCVARAHKMQILMNNNGYDEKQFVGNLKAQT--GGCCVA 79
DB 39 SSGQTCSNQCMSSLCYNGGLYSNTYN-----NNNQYDMNSQ--YGNQMGTTGGSTYCT- 90
OY 80 WSYHVAIVSYKNASGVTEKRIIDPSLFS-----SGPV-----TDTAMRACVN 123
DB 91 -----TSSCRSEFTCMNRCQSSYNNMYNSQYSNSGSGTCTRYSTDCTSGQMCN 142
OY 124 TSCGSASVSYANTAGNVTYRSPNSY-LYDNNL-INTNCV 162
DB 143 GVC---VAQYGTSYNNAMYSSTGNYCTYDWTGCLNGRCV 179

RESULT 14
O7Y235_CRYPV PRELIMINARY; PRT; 405 AA.
ID O7Y235_CRYPV
AC O7Y235;
DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2003, sequence version 1.
DE Hypothetical protein.
GN ORFName=IMB.718;
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxId=5807;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bankier A.T., Spriggs H.F., Fartmann B., Konfortov B.A., Madera M.,
RA Vogel C., Teichmann S.A., Ivens A., Dear P.H.;
RT "Integrated mapping, chromosomal sequencing and sequence analysis of
   Cryptosporidium parvum.";
RL Genome Res. 0:0-0(2003).
CC -----
```

```
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs license
CC -----
DR EMBL, BX538353; CAD98294.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 405 AA; 46655 MW; 4FC406FA937929EF CRC64;

Query Match
Best Local Similarity 28.8%; Pred. No. 7.1;
Matches 32; Conservative 18; Mismatches 47; Indels 14; Gaps 5;

OY 66 YGNLKASTGTCVANSYHVAIVSYKNASGVTEKRIIDPSLFSGGPVDTAMRACVNTS 125
DB 45 YGSLH-----FHRENFCRIKEKGLSGLEKSLID-FIDPLSHADKGTITLIANNHS 96
OY 126 CGSASVSYANTAGNVTYRSPNSY-----YDNNLINTN-C-VLTKFSILS 170
DB 97 DFSQDLGFQMQLTGSVYTRGSDQVLLAQANDSSIRNPECEVTTEFALIA 147

RESULT 15
O7JKS6 CAEEL PRELIMINARY; PRT; 927 AA.
ID O7JKS6 CAEEL
AC O7JKS6;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DE 07-FEB-2006, entry version 19.
GN Hypothetical protein.
OS ORFName=T01D3.6; T01D3.6a;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
   investigating biology.";
RL Science 282:2012-2018(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs license
CC -----
DR EMBL, Z81527; CAB04279.1; -; Genomic_DNA.
DR EMBL, Z81110; CAB04279.1; JOINED; Genomic_DNA.
DR EMBL, Z81110; CAB03262.1; -; Genomic_DNA.
DR EMBL, Z81527; CAB03262.1; JOINED; Genomic_DNA.
DR Ensembl, T01D3.6; Caenorhabditis elegans.
DR Wormbase, WBGene0011330; T01D3.6.
DR WormPep, T01D3.6a; CE12964.
DR GO, GO:0005509; F:calcium ion binding; IEA.
DR InterPro, IPR000152; Asx_hydroxyl_S.
DR InterPro, IPR006210; EGF.
DR InterPro, IPR000742; EGF_3.
DR InterPro, IPR001881; EGF_Ca_bd.
DR InterPro, IPR013091; EGF_Ca_bd.2.
DR InterPro, IPR013032; EGF_like_reg.
DR InterPro, IPR002181; Fibrinogen_C.
DR InterPro, IPR002919; Prot_inh_CR_TTL.
DR InterPro, IPR001846; VWF_D.
DR Pfam, PF07645; EGF_CA; 1.
DR Pfam, PF00147; Fibrinogen_C; 1.
DR Pfam, PF01826; TTL; 1.
DR Pfam, PF00094; VWD; 1.
DR SMART, SM00181; EGF; 3.
DR SMART, SM00179; EGF_CA; 1.
DR SMART, SM00186; FBG; 1.
DR SMART, SM00216; VWD; 1.
DR PROSITE, PS00010; ASX_HYDROXYL; 1.
DR PROSITE, PS00022; EGF_1; UNKNOWN_2.
DR PROSITE, PS01186; EGF_2; 2.
```

DR PROSITE: PS50026; EGF 3; 3.  
 DR PROSITE: PS01187; EGF CA; 1.  
 KM Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 927 AA; 103414 MW; 71AEF0A61FC2B266 CRC64;

Query Match 9.0%; Score 89.5; DB 2; Length 927;

Best Local Similarity 20.1%; Pred. No. 21; Mismatches 68; Indels 149; Gaps 16;  
 Matches 61; Conservative 25;

QY 4 VLPDVATLNSLFNQI--KNQSCGISTASPICITFRYPVDGCTARAHKMQI-----LMN 55  
 DB 130 VCPDYASGRFCQNEIKCKNKGKNA-----DCYVANHQLNLCIKPGYTARR 176  
 QY 56 NGYDCEKQ-----FVYGNLKAS-TGTCVAMS----- 81  
 DB 177 NGRDCDMKVQACMSGDPHYVTYDGLRFYQGTCPYVFSQPCTTLPAPYLMYSVRAXNEL 236  
 QY 82 ----YHVA---ILVSYKN-----ASGVTEKRIIDPSLFS-----S 109  
 DB 237 PKGYHISQVSEVEVDLHNLITHVDGRSKTALVNGV--QVLTPWYFPNKNTWTVVRPS 293  
 QY 110 GP-----VTDTAMFNACY-----NTSCGSA-----SVSSYANTAGNV 141  
 DB 294 GSTFTIENDQGVVFTFTYNSLCVQVPDIPFENGATTLGSLAGNIDGKKLDVNVKNGSV 353  
 QY 142 Y---YRSPSN-----SYLYDNNLI-----NTNCVLTKEPSLIS 170  
 DB 354 LAIKSSROPENNNNHADFMKTEDTWITDKFILRPGQENCINGOTLDNNTNCVSTSIslaQ 413  
 QY 171 GCS 173  
 DB 414 SCA 416

Search completed: September 16, 2006, 06:23:20  
 Job time : 302 secs